

-299-

Identities = 50/58 (86%), Positives = 56/58 (96%)

Query: 137 LHKIELGCDYDNKQSQAVARKLGFTLEANIRDRRDAQGKRCGDMRFGLLRSEWEKKRR 194
 LHKIELGCDYDNKQSQAVARKLGFTLEAN RDR+D QG+RCGDMRFGLLRSEWE++++
 5 Sbjct: 1 LHKIELGCDYDNKQSQAVARKLGFTLEANARDRKDVQGRRCGDMRFGLLRSEWEQKQ 58

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 230

- 10 A DNA sequence (GBSx0244) was identified in *S.agalactiae* <SEQ ID 733> which encodes the amino acid sequence <SEQ ID 734>. This protein is predicted to be ribosomal-protein-alanine N-acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4066(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9599> which encodes amino acid sequence <SEQ ID 9600> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:BAB04418 GB:AP001509 ribosomal-protein-alanine
 N-acetyltransferase [Bacillus halodurans]
 Identities = 63/185 (34%), Positives = 95/185 (51%), Gaps = 11/185 (5%)

Query: 53 KALPKLETDRILRQRTVGDVPAMFDYVCLVEEVAYPAGLSPIASLEDEYDYFENRYQNL 112
 K P LET RLILR+ T D ++ Y+ +EV GL P +LED E +Y+++
 30 Sbjct: 6 KRFPILKRLILRKITDDARSILSYLSDKEVMKYFGLEPFQTLEDALG--EIAWYESI 63

Query: 113 EKAKLPSGYGITVKGSDRIIGSCAFN-----HRHEDDVFEICYLLHPDYWGHGYMTEAVA 167
 + +GIT+KG D +IGSC F+ H + FE+ L YWG G +EA+
 35 Sbjct: 64 LHEQTGIRWGITLKGQDEVIGSCGFHQWVPKHHRAEIGFELSKL----YWGQGIASEAIR 119

Query: 168 ALIEVGFTLLNLHKIEIRCYDNKQSRRAEKLGFTEATIRDRKDNQDNRCVNLIYGLL 227
 A+I+ GF L L +I+ N S+R+ EK GF E +R + +Y LL
 Sbjct: 120 AVIQYGFHEHLELQRIQALIEPPNIPSQLVEKQGFISEGLLRSEYETCGKFDDLYMYSLL 179

40 Query: 228 RSEWE 232
 + +++
 Sbjct: 180 KRDFD 184

There is also homology to SEQ ID 732:

45 Identities = 39/54 (72%), Positives = 44/54 (81%)

Query: 179 LHKIEIRCYDNKQSRRAEKLGFTEATIRDRKDNQDNRCVNLIYGLLRSEWE 232
 LHKIE+ CYDNKQS+ VA KLGFTEA RDRKD Q RC ++ +GLLRSEWE
 50 Sbjct: 1 LHKIELGCDYDNKQSQAVARKLGFTLEANARDRKDVQGRRCGDMRFGLLRSEWE 54

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-300-

Example 231

A DNA sequence (GBSx0245) was identified in *S.agalactiae* <SEQ ID 735> which encodes the amino acid sequence <SEQ ID 736>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 51
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 232

A DNA sequence (GBSx0246) was identified in *S.agalactiae* <SEQ ID 737> which encodes the amino acid sequence <SEQ ID 738>. Analysis of this protein sequence reveals the following:

```

20  Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9597> which encodes amino acid sequence <SEQ ID 9598> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 739> which encodes the amino acid sequence <SEQ ID 740>. Analysis of this protein sequence reveals the following:

```

   Possible site: 38
   >>> Seems to have no N-terminal signal sequence

35  ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3293(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 An alignment of the GAS and GBS proteins is shown below:

```

   Identities = 24/55 (43%), Positives = 38/55 (68%)

   Query: 56  LLEGLTANKQDVLKEAGLVSLFAKVSADVLALKGIGPAAIKQLVDNGVVFQK 110
   ++ G+ ++ + L  G+ S +AF + +E D+LALKGIGPA +K+LV+NG  F K
45  Sbjct: 77  VVAGIRSDLVETLYAEGIHSAQAFKEWTEKDLLALKGIGPATVKKLVENGASFQK 131

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 233

A DNA sequence (GBSx0247) was identified in *S.agalactiae* <SEQ ID 741> which encodes the amino acid sequence <SEQ ID 742>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 743> which encodes the amino acid sequence <SEQ ID 744>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 27
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2536(Affirmative) < succ>
20  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25  Identities = 57/84 (67%), Positives = 73/84 (86%)

Query: 1  MSYEQEFLKDFEEWLQSQISINQMAMDSAKKVLEEDKDERAADAYIRYESKLDAYRFLQG 60
        MSYE+EFLKDFE+W+++QI +NQ+AM ++++V +ED DERA DA+IRYESKLDAY FL G
Sbjct: 1  MSYEKEFLKDFEDWVKTQIQVNQLAMATSQEV AQEDGDERAKDAFIRYESKLDAYEFLLG 60

30  Query: 61 KFNMYHNQKSFHDLPLDGLFGQRHY 84
        KF+NY N K+FHD+PD LFG RHY
Sbjct: 61 KPDNYKNGKAFHDIPDELFGARHY 84

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 234

A DNA sequence (GBSx0248) was identified in *S.agalactiae* <SEQ ID 745> which encodes the amino acid sequence <SEQ ID 746>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

```

40  Possible site: 31
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
45  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 747> which encodes the amino acid sequence <SEQ ID 748>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 35
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>

```

-302-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5 Identities = 26/60 (43%), Positives = 37/60 (61%)
 Query: 23 LKNERCPHPKLINVLERKLEIILGDQKHILEKDSLISLSPQETHHLRAIENSKFLQIELD 82
 + E P K+I VLE +L L DQK +L ++SLI++ Q+ HHL A + K LQ+ LD
 10 Sbjct: 42 ISQETSPRKVILVLEGQLIFDLEDQKQVLTQESLIAIPAQKVHLEAKTDCKLLQVLLD 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 235

15 A DNA sequence (GBSx0249) was identified in *S.agalactiae* <SEQ ID 749> which encodes the amino acid sequence <SEQ ID 750>. This protein is predicted to be integrase (codV). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3842 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 236

30 A DNA sequence (GBSx0250) was identified in *S.agalactiae* <SEQ ID 751> which encodes the amino acid sequence <SEQ ID 752>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> May be a lipoprotein

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 752 (GBS128) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 5; MW 15kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 4; 2 bands).

45 The GBS128-GST fusion product was purified (Figure 198, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 288), which confirmed that the protein is immunoaccessible on GBS bacteria.

-303-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 237

A DNA sequence (GBSx0251) was identified in *S.galactiae* <SEQ ID 753> which encodes the amino acid sequence <SEQ ID 754>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2940(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 755> which encodes the amino acid sequence <SEQ ID 756>. Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.2518(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below:

Identities = 30/90 (33%), Positives = 49/90 (54%), Gaps = 10/90 (11%)

Query: 3 TVAVRVDDQLKDDATELFQSLGLDMSTAVKMFLIQSVKTSIPFEIK-----NKSSV 54
 T+ +RVDD +K A ++ + LG+ MSTA+ MFL Q + T IPF++ N +
30 Sbjct: 15 TLNLRVDDSVKSAADDILKRLGIPMSTAIDMFLNQIILTGGIFFDVSLPEAPQVRNVNDYM 74

Query: 55 SDEEFQNLVETKLGIRVKASDPESVNAFF 84
 S E+ F + + T + K + P+ V F+
35 Sbjct: 75 SQEKFYDKLITSFED--AKTCNPQDVGKFY 102

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 238

A DNA sequence (GBSx0252) was identified in *S.galactiae* <SEQ ID 757> which encodes the amino acid sequence <SEQ ID 758>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.81 Transmembrane 370 - 386 (368 - 388)

----- Final Results -----

45 bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

-304-

A related GBS nucleic acid sequence <SEQ ID 9593> which encodes amino acid sequence <SEQ ID 9594> was also identified. A related GBS nucleic acid sequence <SEQ ID 10773> which encodes amino acid sequence <SEQ ID 10774> was also identified.

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 759> which encodes the amino acid sequence <SEQ ID 760>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL    Likelihood = -4.57    Transmembrane  354 - 370 ( 353 - 371)

10  ----- Final Results -----
        bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  LPXTC motif: 344-348

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 64/277 (23%), Positives = 99/277 (35%), Gaps = 31/277 (11%)

20  Query: 126 SIGNLPDLPKGTTVAFETPVDTATPGDKPAKVVTYPDGSKDTVDVTVKVVDPRTDADKN 185
      ++ +LP  + TT  E PV          + V          + D+ + T    P  A
Sbjct: 121 AVKDLPASTESTTQPVEAPVQETQASADSMVTGdstsvTTDSPEETPSSESPVAPALSE 180

Query: 186 DPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVAFETPVDTATPGDKPAKVVTYPDGSK 245
      PA  Q    E P  S  P    T  A  ETP + A P  P    +  S+
25  Sbjct: 181 APA----QPAESEEPSVAASSEETPS--PSTPAAPETPEEPAAPSPSESEEPSVAAPSE 234

Query: 246 DTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVAFETPVD 305
      +T          P  A  + PA  ++          T  +          P  P  +  +TP
30  Sbjct: 235 ETPSPET----PEEPAAPSPQPAESEESSVAATTSPS-----PSTPAESET--QTTPAV 281

Query: 306 ATPGDKPAKVVTYPDGSKDTVDVTVKVVDPRTDADK-----NDPAGKDQQVNGK 355
      DKP+      P  S  + TV+  + +DK          N  +  + +
35  Sbjct: 282 TKDSKPSGAAEK-PAASSLVSEQT VQPTSKRSSDKKEEQQSYPNRSLSRQVRAHES 340

Query: 356 GNKLPATGENATPFFNVVALTIMSSVGLLSVSKKKED 392
      G  LP+TGE A P F +  +T+MS  G L V+K++++
Sbjct: 341 GKYPSTGEKAQPLF-IAITMTLSLFGSLLVTKRQKE 376

```

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 239

A DNA sequence (GBSx0253) was identified in *S.agalactiae* <SEQ ID 761> which encodes the amino acid sequence <SEQ ID 762>. This protein is predicted to be surface protein Rib. Analysis of this protein
45 sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

50  ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-305-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 240

5 A DNA sequence (GBSx0254) was identified in *S.agalactiae* <SEQ ID 763> which encodes the amino acid sequence <SEQ ID 764>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.06 Transmembrane 39 - 55 (39 - 55)

10

----- Final Results -----
bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15

A related GBS nucleic acid sequence <SEQ ID 9591> which encodes amino acid sequence <SEQ ID 9592> was also identified.

The protein differs significantly from U58333 in several places:

20 Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
T PDG D V+V++ + + DK D K KAED P +G+
Sbjct: 683 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQQVNVGETPKAEDSIGNLPDLPGKTTVA 742

25 Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQQVNV 258
+D T D K T D + +VT K++ PR DADKNDPAGKDQQQVNV
Sbjct: 743 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQQVNV 798

30 Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
T PDG D V+V++ + + DK D K KAED P +G+
Sbjct: 841 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQQVNVGETPKAEDSIGNLPDLPGKTTVA 900

35 Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQQVNV 258
+D T D K T D + +VT K++ PR DADKNDPAGKDQQQVNV
Sbjct: 901 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQQVNV 956

40 Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
T PDG D V+V++ + + DK D K KAED P +G+
Sbjct: 288 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQQVNVGETPKAEDSIGNLPDLPGKTTVA 347

45 Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQQVNV 258
+D T D K T D + +VT K++ PR DADKNDPAGKDQQQVNV
Sbjct: 348 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQQVNV 403

50 Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
T PDG D V+V++ + + DK D K KAED P +G+
Sbjct: 604 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQQVNVGETPKAEDSIGNLPDLPGKTTVA 663

55 Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQQVNV 258
+D T D K T D + +VT K++ PR DADKNDPAGKDQQQVNV
Sbjct: 664 FETPVDTA-TPGDKPAKVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQQVNV 719

60 Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
T PDG D V+V++ + + DK D K KAED P +G+

-306-

Sbjct: 920 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPGKTTVA 979

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPDADKNDPAGKDQQVNV 258
 +D T D K T D + +VT K++ PR DADKNDPAGKDQQVNV

5 Sbjct: 980 FETPVDTA-TPGDKPAKVVVITYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 1035

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 241

A DNA sequence (GBSx0255) was identified in *S.agalactiae* <SEQ ID 765> which encodes the amino acid sequence <SEQ ID 766>. This protein is predicted to be ara-C-like activator. Analysis of this protein sequence reveals the following:

15 Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 25)

----- Final Results -----
 20 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9589> which encodes amino acid sequence <SEQ ID 9590> was also identified.

25 There is homology to SEQ ID 460.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 242

30 A DNA sequence (GBSx0256) was identified in *S.agalactiae* <SEQ ID 767> which encodes the amino acid sequence <SEQ ID 768>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.1200(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9587> which encodes amino acid sequence <SEQ ID 9588> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 769> which encodes the amino acid sequence <SEQ ID 770>. Analysis of this protein sequence reveals the following:

45 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0679(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 135/176 (76%), Positives = 161/176 (90%)

```

5   Query: 1   MSYMKDRQIQKTKVAIYNFISLLQENDYSKITVQDVIGLANVGRSTFYSHYESKEVLL 60
      +S M KDRQI+KTK AIY+AFI+LLQ+ +YSKITV+D+I LANVGRSTFY+HYESKE+LL
      Sbjct: 1   VSDMTKDRQIKKTKTAIYSAFIALQKKEYSKITVRDMITLANVGRSTFYAHYESKEMLL 60

10  Query: 61   KELCEDLFHHLFKQGRDVTPEEYLVHILKHFEQNQDSIATLLLSDDPYFLLRFRSELEHD 120
      KELCE+LFHHLF+Q R+VTFE+YLVHILKHFEQN+DSIATLLLS+DPYFLLRF++ELEHD
      Sbjct: 61   KELCEELFHHLFRQKRNVTFEDYLVHILKHFEQNKDSIATLLLSNDPYFLLRFKNELEHD 120

      Query: 121 VYPRLREEYITKVDIPEDFLKQFLSSFIETLKWHLHQRQKMTVEDLLKYLLTMVE 176
      VYP LR +YI K IPE FLKQF+LSSFIETLKWHLHQRQ+M+ +LLKYLL +++
15  Sbjct: 121 VYPNLRCKYIDKTTIPEVFLKQFVLSSFIETLKWHLHQRQMSANELLYLELIK 176

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 243

20 A DNA sequence (GBSx0257) was identified in *S.agalactiae* <SEQ ID 771> which encodes the amino acid sequence <SEQ ID 772>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3573(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 244

35 A DNA sequence (GBSx0258) was identified in *S.agalactiae* <SEQ ID 773> which encodes the amino acid sequence <SEQ ID 774>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.

40	INTEGRAL	Likelihood = -10.19	Transmembrane	112 - 128 (107 - 131)
	INTEGRAL	Likelihood = -8.07	Transmembrane	77 - 93 (71 - 97)
	INTEGRAL	Likelihood = -6.10	Transmembrane	144 - 160 (138 - 165)
	INTEGRAL	Likelihood = -3.03	Transmembrane	165 - 181 (164 - 182)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.5076(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 775> which encodes the amino acid sequence <SEQ ID 776>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

	INTEGRAL	Likelihood = -9.13	Transmembrane	112 - 128 (107 - 130)
--	----------	--------------------	---------------	------------------------

-308-

```

INTEGRAL    Likelihood = -5.89    Transmembrane  144 - 160 ( 138 - 163)
INTEGRAL    Likelihood = -5.47    Transmembrane   7 - 23 (   6 - 29)
INTEGRAL    Likelihood = -3.50    Transmembrane  77 - 93 (  74 - 94)
INTEGRAL    Likelihood = -2.07    Transmembrane 166 - 182 ( 165 - 183)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 212/287 (73%), Positives = 245/287 (84%)

```

Query: 1   MTSNKKVAIAFIILNISFSVLEFIFGSLFFSGAILADAVHDFGDAIAIGISATLEKKSKKD 60
          M ++KKV I FILN+SFS++EFIFG+LFFSGAILADAVHDFGDAIAIGISA LE+K+ K
Sbjct: 1   MPASKKVITIIFILNLSFSLIEFIFGTLFFSGAILADAVHDFGDAIAIGISAILEKAVKK 60

Query: 61  EDTIFSLGYKRFSLLGALITSLILISGSILVMIEIPKLWHPTPVNYHGMFILAVIAIII 120
          E   FSLGYKRFSLLGAL T+LILISGS+LVMIE IPKLWHPT VNY GMF+LA+ AIII
Sbjct: 61  ESPNFSGLGYKRFSLLGALTNTNLLISGSLLVMETIPKLWHPTIVNYDGMFVLAIFAI 120

Query: 121 NGLASFILHSGQSKHEEILSLHFLEDILGWLAIIVISLILNWKPLYILDPLLSVAISTFI 180
          NG ASFI+HS Q+K+EEILSLHFLEDILGWLAI++SLIL WKP YILDPLLS+AI++FI
Sbjct: 121 NGFASFIIHSNQTKNEEILSLHFLEDILGWLAIILSLILKWKPYILDPLLSIAIASFI 180

Query: 181 LSKALPKLLSTLKLFLDGVPSIDYAAALHDELKGLSQVRSINQLNIWSMDGIDNRAIHC 240
          LSKALPKL++T +FLDGVPSIDY LH EL L + S+NQLN+WSMDGID+RA IHC
Sbjct: 181 LSKALPKLVATANIFLDGVPSIDYCTLHHELSQLPHIVSVNQLNVWSMDGIDHRATIHC 240

Query: 241 CLNQLISEKDCKRAIRTICQHYKINDVTVEIDYSLREHQNHCKPLKN 287
          CL + +EK CK++IR ICQ Y IN VTVEID SL EHQ+HC L +
Sbjct: 241 CLRESTTEKHCKKSIRLICQRYNINSVTVEIDTSLNEHQHHCSSLSS 287

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 245

A DNA sequence (GBSx0259) was identified in *S.agalactiae* <SEQ ID 777> which encodes the amino acid sequence <SEQ ID 778>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -1.22    Transmembrane  221 - 237 ( 221 - 237)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 246

A DNA sequence (GBSx0260) was identified in *S.agalactiae* <SEQ ID 781> which encodes the amino acid sequence <SEQ ID 782>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -2.50    Transmembrane   2 - 18 (   1 - 18)

```

----- Final Results -----

5 bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 10 vaccines or diagnostics.

Example 247

A DNA sequence (GBSx0261) was identified in *S.agalactiae* <SEQ ID 783> which encodes the amino acid
 sequence <SEQ ID 784>. This protein is predicted to be dehydrogenase (Zn-dependent). Analysis of this
 protein sequence reveals the following:

15 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.77 Transmembrane 171 - 187 (170 - 187)

 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:AAG20655 GB:AE005134 alcohol dehydrogenase; Adh2 [Halobacterium
 sp. NRC-1]
 Identities = 169/348 (48%), Positives = 232/348 (66%), Gaps = 9/348 (2%)

 30 Query: 1 MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWYRGISKRESGSFAGHE 60
 M+ A + PG++ + + PKP IE DAVI++ VCGSDLW+YRG S RE+GS GHE
 Sbjct: 1 MRAAVYQGPGEIAVEEVKPKDIESPEDAVIRVTHTAFCGSDLWFYRGSDREAGSRVGHE 60

 Query: 61 AIGIVEEVGKTKVTDVSKGDFVIVPFFTHGCGQCPCKAGFDGNCNTHQA---AKNVGYQGO 117
 +GIVEEVG VT V+ GD VI PF CG+C C+ G +C ++ N G QG+
 35 Sbjct: 61 PMGIVEEVGDDVTSVAPGDRVIAPFAISCGECEFRCRQGLYTSCVEDESWGSEANGGGQGE 120

 Query: 118 YLRYTNANWALVKIPGQPSDYDNETLNSLLTLSDVMATGYHAAATAEVKEGDTVVVMGDD 177
 Y++ A+ LV++P + +D D + L SLL L+DVM TG+HAA +A V EGDV VV+GDG
 40 Sbjct: 121 YVKCFPADGTLVRVEDRYAD-DEVDLESLLPLTDVMGTGHHAASAGVGEEDTAVVVGDD 179

 Query: 178 AVGLCGVIAAKMLGANRIIAMSRLKDRQELALTFGATDIVEERGDEAVKRVLDLTNQAGA 237
 AVGLCGV+AA+ LGA RIIAM H+DR ELA FGATD + RGD+A++R DLT+ GA
 Sbjct: 180 AVGLCGVLAQRLGAERI IAMGHEDRLAELAAFGATDTISARGDDAIERARDLTH-GGA 238

 45 Query: 238 DAVLECVGTEQSVDTATQIARPGAVIGRVGIP---QNPDMNTNNLFWKNIGLRGGIASVT 294
 + V+ECVG ++D+A IARPG +G VG+P ++ ++ +F NI +RGG+A V
 Sbjct: 239 NHVMECVGAASAMDSAIATARPGGTGTVGYGVVPYGVEDGGLDVFTMFSDNITIRGGVAPVR 298

 Query: 295 TFDKSVLLDAVLTHKINPGLVFTKSFVLDLDDIQKAYEAMDKRAIKSLV 342
 + + ++ D VL ++P +FTK+ LD + + Y AMD R+AIK LV
 50 Sbjct: 299 AYAEELMAD-VLQGTLDPSIFTKTKTVLDLGDVPEGYAAMDREAIKVLV 345

There is also homology to SEQ ID 786.

A related sequence was also identified in GAS <SEQ ID 9145> which encodes the amino acid sequence
 55 <SEQ ID 9146>. Analysis of this protein sequence reveals the following:

Possible site: 23

-310-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.41 Transmembrane 170 - 186

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 121/353 (34%), Positives = 182/353 (51%), Gaps = 16/353 (4%)

Query: 1 MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRG-ISKRESGSFAGH 59
 MK AT++ G + + D PKPVI + TDA++++V+ +CG+DL G + + G+ GH

15 Sbjct: 15 MKAATYLSTGNLQLIDKPKPVIKPTDAIVQLVKTTCGTDLHLGGDVPACKEGTILGH 74

Query: 60 EAIGIVEEVGTVKTDVSKGDFVIVPPTHGCGQCPCKAGFDGNCNTHQAAKN---VGYYQG 116
 E IGIV+EVG VT+ GD VI+ C C CK G +C + G Q

20 Sbjct: 75 EGIGIVKEVGDAVTNFKIGDKVIISCVTSCHTCYCKRGLSSHCDGGWILGHLINGTQA 134

Query: 117 QYLRVTNANWALVKIPGQPSDYDNETLNSLLTSLSDVMATGYH-AAATAEVKEGDTVVMVG 175
 +Y+ +A+ +L P D +L+ LSD++ T Y + VK GD V ++G

Sbjct: 135 EYVHIFPHADGSLYHAPDTIDD-----EALVMLS DILP TSYEIGVLP SHVKPGDNVCIVG 188

25 Query: 176 DGAVGLCGVIAAKMLGANRIIAMS RHKDRQELALTFGATDIVEERGDEAVKRVL-DLTNQ 234
 G VGL ++ + II + ++R E A TFGAT + E VK ++ D+TN

Sbjct: 189 AGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHHTICSGSSEEVKAIIDITNG 248

Query: 235 AGADAVLECVGTEQSVDTATQIARPGAVIGRVGIPQNP-DMNTNNLFWKNIGLRGGIASV 293
 G D +ECVG + D +I G I VG+ P D N + L+ KNI L G+ +

30 Sbjct: 249 RGVDISMECVGYPATFDICQKIISVGGHIANVG VHGKPVDFNLDELWIKNITLNTGLVNA 308

Query: 294 TTFDKSVLLDAVLTHKINPGLVFTKSFVLD DDIQKAYEAMDKRDAIKSL-VIVD 345
 T + +LL+ + T KI+ + T F L +++KAYE A +L VI+D

35 Sbjct: 309 NITE--MLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNALKVIID 359

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 248

40 A DNA sequence (GBSx0262) was identified in *S.agalactiae* <SEQ ID 787> which encodes the amino acid sequence <SEQ ID 788>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2169(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD36075 GB:AE001762 hypothetical protein [Thermotoga maritima]
 Identities = 55/128 (42%), Positives = 72/128 (55%), Gaps = 8/128 (6%)

Query: 8 IFPKGEKNPYGEFFIGQSYLAALAKSPDG--NVSVGNVTFEAGCRNNWHVHLDGYQILLV 65
 IF +G K +FF G ++ L +G N V +V FE G R +WH H G QIL+V

55 Sbjct: 5 IFERGSKGS-SDFFTGNVWVKMLVTDENGVFNTQVYDVVFEPGARTHWHSHPGG-QILLV 62

Query: 66 TEGSGWYQEEGKEAVSLKPGDVIVTDKGVVRHWHGAKKDSEFAHIAITA----GKSEFYEA 121
 T G G+YQE GK A LK GDV+ V HWHGA D E HI I+ G +E+ +

60 Sbjct: 63 TRKGFGYQERGKPARILKKGDVVEIPPNVHWHGAAPDEELVHIGISTQVHLGPAEWLGS 122

-311-

Query: 122 VSDEEYSR 129
 V++EEY +
 Sbjct: 123 VTEEEYRK 130

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 249

- 10 A DNA sequence (GBSx0263) was identified in *S.agalactiae* <SEQ ID 789> which encodes the amino acid sequence <SEQ ID 790>. This protein is predicted to be gamma-carboxymuconolactone decarboxylase. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4089(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA20070 GB:AL031155 3-oxoadipate enol-lactone
 hydrolase/4-carboxymuconolactone decarboxylase
 [Streptomyces coelicolor A3(2)]
 Identities = 33/93 (35%), Positives = 59/93 (62%), Gaps = 1/93 (1%)

- 25 Query: 11 QLEEFAPFARYNDDILFGEVWAKEDHLTDKTRSIITISALISGGNLEQLEHHLQFAKQN 70
 Q +EF+ +F + +GE+W + L ++RS +T++AL++GG+L++L HL+ A +N
 Sbjct: 349 QADEFSGDFQEFTRYAWGEIWDPRG-LDRRSRSCVTLTALVAGGHLDELAPHLRAALRN 407
- 30 Query: 71 GVTKEEIADIITHLAFYVGWPKAWSAFNKAKEI 103
 G+T EI +++ A Y G P A AF A+++
 Sbjct: 408 GLTPGEIKEVLLQAAVYCGVPAANGAFRVAQOV 440

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 250

A DNA sequence (GBSx0265) was identified in *S.agalactiae* <SEQ ID 791> which encodes the amino acid sequence <SEQ ID 792>. Analysis of this protein sequence reveals the following:

- 40 possible site: 44
 >>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5529(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 251

A DNA sequence (GBSx0266) was identified in *S.agalactiae* <SEQ ID 793> which encodes the amino acid sequence <SEQ ID 794>. This protein is predicted to be probable transcriptional regulator. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9585> which encodes amino acid sequence <SEQ ID 9586> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG08263 GB:AE004901 probable transcriptional regulator
[Pseudomonas aeruginosa]
Identities = 36/148 (24%), Positives = 68/148 (45%), Gaps = 22/148 (14%)
Query: 5 QIVEKPFAMILAG-----VTLENVKSNQEGIQQAIGICKTQPDFRFD 45
+IVE+PA + G + E+ + + + + GIC QP+ F
Sbjct: 123 RIVERPAFVSVVGMFYFGSAPGDTIGQLWERFIPREHEIAGKHDPEVSYGICAAQQPNGEFH 182
Query: 46 YSATYQVETSVQAPKGLIIRIPSATYAVISVKGPMPSSLQETWRKIIQGGFFQENNLKPA 105
Y A ++V+ P+G+ ++P+ YAV + KG P + E+++ I E L+P
Sbjct: 183 YVAGFEVQEGWPVPEGMVRFQVPAQKYAVFTHKGTAP-QIAESFQAIYSHLLAERGLEPK 241
Query: 106 NSPNLEIYSSQH--PQDTDYQMEIWLAI 131
+ E Y + P D + Q++++ I
Sbjct: 242 AGVDFEYYDQRFGRGLDPNSQVDLYIPI 269

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 252

A DNA sequence (GBSx0267) was identified in *S.agalactiae* <SEQ ID 795> which encodes the amino acid sequence <SEQ ID 796>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0887 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB84919 GB:AE000825 conserved protein [Methanothermobacter
thermoautotrophicus]
Identities = 42/130 (32%), Positives = 71/130 (54%), Gaps = 3/130 (2%)

-313-

Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
 M+T EM + I +L VAT D +G PN+ P R D++T + +N +T N+ +N
 Sbjct: 1 MMTPEMMDAIEKELVVFVATADEEGTPNVVPIGFARPLDERTILIADNYMKKTIRNLHENP 60

5 Query: 61 KIEIAFVDRELLGYRFGVTAEIQTEGTYEAAKKWAEGRMG--VPKAVGIIHVERIFNL 118
 +I + R Y+F GT EI G Y++ +WA+ M PK+ ++ VE I+++
 Sbjct: 61 RIAL-IPQNARECPYQPKGTVEIFKSGKYFDMVVEWAQNVMTLEPKSAILMTVEEITYSV 119

10 Query: 119 QSGANAGKEI 128
 + G AG+++
 Sbjct: 120 KPGPEAGEKV 129

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 797> which encodes the amino acid sequence <SEQ ID 798>. Analysis of this protein sequence reveals the following:

15 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0789(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 123/128 (96%), Positives = 127/128 (99%)

25 Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
 MITQEMK++IN+QLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG
 Sbjct: 1 MITQEMKDLINNQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60

30 Query: 61 KIEIAFVDRELLGYRFGVTAEIQTEGTYEAAKKWAEGRMGVPAKVGIIHVERIFNLQS 120
 KIEIAFVDRELLGYRFGVTAEIQTEG YYEAAKKWA+GRMGVPAKVGIIHVERIFNLQS
 Sbjct: 61 KIEIAFVDRELLGYRFGVTAEIQTEGAYYEAKKWAQGRMGVPAKVGIIHVERIFNLQS 120

35 Query: 121 GANAGKEI 128
 GANAGKEI
 Sbjct: 121 GANAGKEI 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 253

A DNA sequence (GBSx0268) was identified in *S.agalactiae* <SEQ ID 799> which encodes the amino acid sequence <SEQ ID 800>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -5.47 Transmembrane 1028 -1044 (1027 -1048)

----- Final Results -----
 bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

!GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 55 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 >GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides

forsythus]
Identities = 143/566 (25%), Positives = 243/566 (42%), Gaps = 52/566 (9%)

5 Query: 95 VPKAKPEVTQEASNSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVG---F 150
+P + + + A + + +P TA + L T + T +G F
Sbjct: 120 IPNSVTITIGEWAFKGCGLKSITLPSNLTAIGQSALSGCTGLTSITIPNSVTITIGEWAF 179

10 Query: 151 SKSGINKLSQTSHLVLP SHA--DGTQLTQVASFAPDPKKTATIAEYTSRLGENGKPSRL 208
SG+ ++ + L +A LT + PD T I E + G +G S
Sbjct: 180 GCSGLTSITFPNSLTAIGESAFYGCGLTSIT---LPDALTTIGESAFK-GCSGLKSIT 234

15 Query: 209 DIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVLNLPESLETISDYAFA 268
+ I E ++ LT +T+P+ +IG+ AF + + P SL TI + AF
Sbjct: 235 FPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCGLKSITFPNSLTTIGESAFY 294

20 Query: 269 HM-SLKQVKLPDNLKVGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKL 327
+ SL + +P+++ IG AF+ + LP L + ERAF + + T + + +
Sbjct: 295 NCGSLTSITIPNSVTITIGRSAFYGCGLKSITLPDGLTTIEERAFYNCGLVLSITIPNSV 354

25 Query: 328 KVIGESAFQD-NLNRNVLMPDGLKIESEAFIGNPGDEHYNNQVLRTRTGONPHQLATE 386
IGE++F + L+++ LPDGL IE AF N L + T N E
Sbjct: 355 ATIGESAFYGCGLKSITLPDGLTTIEWGAFY-----NCGALTSITIPNSVSTIGE 405

30 Query: 387 NTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFS---NKGLQKVRNKNLEIPKQH 443
+ + +L T D ++ D +++ ++G G + V K ++ K+
Sbjct: 406 SAFYGCGL-ALKDVTVAWDTPIDIQRD-VFRELTLSGIRLHVPAKKTIVYEAK--DVWKEF 461

35 Query: 444 NGITITEIGDNAFRNVDFQSKTLRKVDLEEIKLPSTIRKIGAFQSNLKSFEASEDL 503
N + + G + ND +KTL + P T + + FA ++ L
Sbjct: 462 NIVEDDDFGGLQW-NYDAATKLTITN---PTPDTPKMPNFATPNDQLW----- 507

40 Query: 504 EIKEGAFMNNRIGTLDLKDLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHL 562
GAF I + + D + +GD AF + + +I LP+SV IG+SAF L
Sbjct: 508 ----GAFQKE-IQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRS 562

45 Query: 563 MFIGNKVKITIGEMAFLSNKLESVNLSEQQLKTIEVQAFS-DNALSEVVLPNLTIRE 621
+ + + V TIGE AF + LE +++ K + I + F +L+ + LP L I ++
Sbjct: 563 LTLPGVNTTIGEKAFY-DCLBLTSITIPKSVTAIGQETTFHYCVSLTSLTLPDALTAIGKK 621

50 Query: 622 AF-KRNHLKEVKGSSTLSQITFNADF 646
AF N L V +++ I NADF
Sbjct: 622 AFYSCNALTSVTFPKSITTIGENADF 647
Identities = 109/407 (26%), Positives = 175/407 (42%), Gaps = 48/407 (11%)

55 Query: 222 FNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVLNLPESLETISDYAFAHMS-LKQVKLPDN 280
F+ LT +T+PN +IG AF + + +P S+ TI ++AF S LK + LP++
Sbjct: 87 FSDCALTSVTLPSNLTAIGDHAFKGCGLTSITIPNSVTITIGEWAFKGCGLKSITLPNS 146

60 Query: 281 LKVGELAPFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGESAFQD-NN 339
L IG+ A + +P + + E AF T + L IGE++F
Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTITIGEWAFKGCGLTSITFPNSLTAIGESAFYGC 206

65 Query: 340 LRNVMLPDGLKIESEAFIGNPGDEHYNNQVLRTRTGONPHQLATENTYVNPDKSLWRA 399
L ++ LPD L I AF G G L++ T N E+ + +
Sbjct: 207 LTSITLPDALTTIGESAFKGCSCG-----LKSITFPNSLTTIGESAFYDCGALTSIT 257

Query: 400 TPDMDYTKWLEEDFTYQKNSVTGFSNKGLOKVRNKNLEIPKQHNIGITITEIGDNAFRNV 459
PD ++T K++ P ++T IG++AF N
Sbjct: 258 LPD-----ALTTIGRSAFYGCGLKSITFPN-----SLTTIGESAFYNC 296

Query: 460 DFQSKTLRKVDLEEIKLPSTIRKIGAFQSNLKSFEASEDLKEIKEGAFMNNRIGT- 517
L I +P+++ IG AF + LKS + L I+E AF N + T
Sbjct: 297 G-----SLTSITIPNSVTITIGRSAFYGCGLKSITLPDGLTTIEERAFYNCGLVLS 347

Query: 518 LDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKITIGEMA 576
+ + + + IG++AF+ + + +I LP+ + I AF GAL + I N V TIGE A
Sbjct: 348 ITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAFYNCALTSITIPNSVSTIGESA 407

-315-

- Query: 577 FLS-NKLESVNLSEQKQLKTIEVQAFSDNALSEVVL--PPNLQTIRE 620
 F L+ V ++ + I+ F + LS + L P +T+ E
 Sbjct: 408 FYGCGALKDVTVAWDTPI-DIQRDVFRELTLGIRLHVPAGKKTVE 453
 Identities = 111/465 (23%), Positives = 185/465 (38%), Gaps = 56/465 (12%)
- 5 Query: 141 VTRGDTLVGFSKSGINKLSQTSHLVLP SHAADGTQLTQVASFAF-----TPDKKT 190
 +T D L +S S + P+ LT + AF PD T
 Sbjct: 210 ITLPDALTTIGESAFKGCGLKSITFPN-----SLTTIGESAFYDCGALTSITLPDAL 263
- 10 Query: 191 AIAEYTSRLGKSGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKN 250
 I ++ G +G S + I E +N LT +TIPN +IG+ AF +
 Sbjct: 264 TIGR-SAFYGCGLKSITFPNSLTITIGESAFYNGSLTITIPNSVTITIGRSAFYGCGL 322
- 15 Query: 251 AENVLPESLETISDYAFAHMS-LKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAE 309
 + LP+ L TI + AF + L + +P+++ IGE AF+ + LP L +
 Sbjct: 323 KSITLPDGLTTIEERAFYNGCVLTITIPNSVATIGESAFYGCGLKSITLPDGLTTIEW 382
- 20 Query: 310 RAFKSNRIQTVEFLGSKLVIGESAFQD--NNLRNVMLP-DGLEKIESEAF-----TGNPG 362
 AF + T + + + IGE++F L++V + D I+ + F +G
 Sbjct: 383 GAFYNGGALTSITIPNSVSTIGESAFYGCGLKDVTVAWDTPIDIQRDVFRELTLGIRL 442
- 25 Query: 363 DEHYNNQVVLRTTRTGQNPQLATEN-----TYVNPDKSLWRATPDMDYTKWLEEDFTY 415
 + V + + ++ Y K+L P D K + +F
 Sbjct: 443 HVPAGKKTVEAKDVWKEFNIVEDDFGGLQWNYDAATKTITITNETPTPKPM-PNFAT 501
- 30 Query: 416 QKNSVTGFSNKGKLVIGESAFQDNNLRNVMLPDGLEKIESAFTGNPGDEHYNNQVVLRT 475
 + + G K +QK+ G +T +GD AF D L+ I
 Sbjct: 502 PNDQLWGAFQKEIQKIT-----IGDGVTSVGDFAFSGCD-----ALKSIT 541
- 35 Query: 476 LPSTIRKIGAFAFQSN-NLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDLIKIGDAAFH 533
 LP ++ IG AF +L+S + + I E AF + + ++ + + IG FH
 Sbjct: 542 LPKSVTTIGQSAFSGCWDLRSLTLPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFH 601
- 40 Query: 534 -INHIYAIPLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAF 577
 + ++ LP+++ IG+ AF AL + + TIGE AF
 Sbjct: 602 YCVSLTSLTLPDALTAIGKKAFYSCNALTSTVTFPKSITTIGENAF 646
 Identities = 98/351 (27%), Positives = 152/351 (42%), Gaps = 53/351 (15%)
- 45 Query: 315 NRIQTVEFLGSKLVIGESAFQDNNLRNVMLPDGLEKIESAFTGNPGDEHYNNQVVLRT 374
 ++IQTV +G + +G +F D L +V LP+ L I AF G G L +
 Sbjct: 68 SKIQVT-IGDGVTSVGNNAFSDCALTSVTLPNSLTAIGDHAFKGCSCG-----LTS 117
- 50 Query: 375 RTGQNPQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFSNKGKLVIGESAFQD 434
 T P+ + T + S ++ NS+T L
 Sbjct: 118 IT--IPNSVTTIGEWAFKGCGLKSIT-----LPNSLTAIGQSALSGCTGL 161
- 55 Query: 435 KNLEIPKQHNIGITITEIGDNF-----RNVDFQSKTLRKVD-----LEEIKLPSTI 480
 ++ IP ++T IG+ AF ++ F + + L I LP +
 Sbjct: 162 TSITIPN-----SVTTIGEWAFKGCGLTSITFPNSLTAIGESAFYGCGLTSITLPDAL 216
- 60 Query: 481 RKIGAFAFQS--NNLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDLIKIGDAAFH-INHI 537
 IG AF+ + LKS L I E AF + + ++ L D L IG +AF+ + +
 Sbjct: 217 TTIGESAFKGCGLKSITFPNSLTITIGESAFYDCGALTSITLPDALTTIGRSAFYGCGL 276
- 65 Query: 538 YAIPLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFSL-NKLESVNLSEQKQLKTI 596
 +I P S+ IG SAF G+L + I N V TIG AF + L+S+ L + L TI
 Sbjct: 277 KSITFPNSLTITIGESAFYNGSLTITIPNSVTITIGRSAFYGCGLKSITLPD--GLTTI 334
- Query: 597 EVQAFSD-NALSEVVLPPNLQTIREAFKR-NHLKEVKGSSLTSLQITFN 645
 E +AF + L+ + +P ++ TI E AF + LK + L+ I + AF
 Sbjct: 335 EERAFYNGCVLTITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAF 385
 Identities = 78/282 (27%), Positives = 123/282 (42%), Gaps = 46/282 (16%)
- Query: 111 NDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP-- 168
 N+AS E+P SK +T VT GD + + + + TS + LP+
 Sbjct: 56 NNAS--EIPWHSLSQSKIQT-----VTIGDGVTSVGNNAFSDCALTS-VTLPNSL 101
- Query: 169 -----HAADG-----TQLTQVASFAFT-----PDKKTAIAEYTSRLGK 203

-316-

HA G +T + +AF P+ TAI + ++ G G
 Sbjct: 102 TAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPNSLTAIGQ-SALSGCTG 160

Query: 204 KPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETIS 263
 S + I E F LT +T PN +IG+ AF + + LP++L TI
 Sbjct: 161 LTSITIPNSVTTIGEWAFKGCGLTSITFPNSLTAIGESAFYGCGLTSITLPDALTTIG 220

Query: 264 DYAFAHMS-LKQVKLEPDNLKIVIGELAFFDNQIGGKLYLPRHLIKLAERAFKS-NRIQTVE 321
 + AF S LK + P++L IGE AF+D + LP L + AF + +++
 Sbjct: 221 ESAFKGCGLKSITFPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCGLKSIT 280

Query: 322 FLGSKLKVIGEEASFQD-NNLRNVMLPDGLEKIESEAF TGNPG 362
 F S L IGE++F + +L ++ +P+ + I AF G G
 Sbjct: 281 FPNS-LTTIGESAFYNGSLTSITIPNSVTTIGRSAFYGCSCG 321

Identities = 43/144 (29%), Positives = 70/144 (47%), Gaps = 4/144 (2%)

Query: 220 EIFNAYQ--LTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFAHM-SLKQVK 276
 +++ A+Q + K+TI +G S+G AF + + LP+S+ TI AF+ L+ +
 Sbjct: 505 QLWGAFOKEIQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRSLT 564

Query: 277 LPDNLKIVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGEEASFQ 336
 LPD + IGE AF+D + +P+ + + + F T L L IG+ +F
 Sbjct: 565 LPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFHVCVSLTSLTLPDALTAIGKKAFY 624

Query: 337 D-NNLRNVMLPDGLEKIESEAF T G 359
 N L +V P + I AF G
 Sbjct: 625 SCNALTSVTFPKSITTIGENAFDG 648

Identities = 43/134 (32%), Positives = 66/134 (49%), Gaps = 12/134 (8%)

Query: 511 MNNRIGTLDLKDCLKIKIGDAAFINHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVK 570
 + ++I T+ + D + +G+ AF + ++ LP S+ IG AF+ L + I N V
 Sbjct: 66 LQSKIQTVTIGDGVTSVGNNAFSDCALTSVTLPNSLTAIGDHAFKGCGLTSITIPNSVT 125

Query: 571 TIGEMAFLS-NKLESVNLSSEQQLKTIEVQAFSD-NALSEVVLPPNLQITREEAFKRNHL 628
 TIGE AF + L+S+ L L I A S L+ + +P ++ TI E AF
 Sbjct: 126 TIGEWAFKGCGLKSITL--PNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAF----- 178

Query: 629 KEVKGSSTLSQITF 642
 G S L+ ITF
 Sbjct: 179 ---FGCSGLTSITF 189

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 801> which encodes the amino acid sequence <SEQ ID 802>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.44 Transmembrane 984 -1000 (984 -1001)

----- Final Results -----
 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 975-979

An alignment of the GAS and GBS proteins is shown below:

Identities = 751/1050 (71%), Positives = 861/1050 (81%), Gaps = 45/1050 (4%)

Query: 3 KKHLKTLALALTTVSVVITYSQEVYGLEREESVKQEQTQSA-SEDDWFEEEDNERKTNVSKE 61
 KKHLKT+AL LTTVSVVTT++QEV+ L +E +KQ Q S+ S D+ E + K +++
 Sbjct: 2 KKHLKTVALTLTTVSVVTHNQEVFLVKEPILKQTQASSSISGADYAESSGSKLKINET 61

Query: 62 NSTVDETVDLFDGNSNNSSSKTESVVSDDPKQVPAKPEVTQEASNSSNDASKVEVPKQ 121
 + VD+TV+DLFSD + K +Q KA E T E+ S++E K+
 Sbjct: 62 SGPVDDTVIDLFDSDKRTTPEKIKDNLAGPREQELKAVTENT-ESEKQITSGSQLEQSKE 120

-317-

Query: 122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP SHAADGTQLTQVAS 181
 + K TS WE DF+T+G+TLVG SKSG+ KLSQT HVLVPS AADGTQL QVAS
 Sbjct: 121 SLSLNKTVPSTSNWEICDFITKGNLTVLGLSKSGVEKLSQTDHLVLP SQAADGTQLIQVAS 180

Query: 182 FAFTPDKKTAIAEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQ 241
 FAFTPDKKTAIAEYTSR GENG+ S+LD+D KEII+EGE+FN+Y L K+TIP GYK IGQ
 Sbjct: 181 FAFTPDKKTAIAEYTSRAGENGEISQLDVGKEIINEGEVFN SYLLKKVTIPTGYKHIGQ 240

Query: 242 DAFVDNKNIAEVLNLPESLETISDYAFAHMSLKQVKLPDNLKVIGELAFFDNQIGGKLYLP 301
 DAFVDNKNIAEVLNLPESLETISDYAFAH++LKQ+ LPDNLK IGE LAFFDNQI GK LP
 Sbjct: 241 DAFVDNKNIAEVLNLPESLETISDYAFALALKQIDLPDNLK AIGELAFFDNQITGKLSLP 300

Query: 302 RHLIKLAERAFKSNRIQTVEFLGSKLVIGEASFQDNLRNVMLEPDGLEKIESEAFNGP 361
 R L++LAERAFKSN I+T+EF G+ LK VIGEASFQDN+L +MLPDGLEKIESEAFNGP
 Sbjct: 301 RQLMRLAERAFKSNHIKTIEPRGNSLVIGEASFQDNDSLQMLPDGLEKIESEAFNGP 360

Query: 362 GDEHYNNQVVLRTGTGQNPQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421
 GD+HYNN+VVL T++G+NP LATENTYVNPDKSLW+ +P++DYTKWLEEDFTYQKNSVT
 Sbjct: 361 GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

Query: 422 GFSNKGLOKVRNKNLEIPKQHNITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIR 481
 GFSNKGLOKV+RNKNLEIPKQHN+TITEIGDNAFRNVDFQ+KTLRKYDLEE+KLPSTIR
 Sbjct: 421 GFSNKGLOKVRNKNLEIPKQHNVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480

Query: 482 KIGAFAPQSNLKSFEASEDLEEIKEGAFMNNRIGTLDLKD KLIKIGDAAFHINHIAIV 541
 KIGAFAPQSNLKSFEAS+DLEEIKEGAFMNNRI TL+LKDKL+ IGDAAFHINHIAIV
 Sbjct: 481 KIGAFAPQSNLKSFEASDDLEEIKEGAFMNNRIETLELKD KLVITIGDAAFHINHIAIV 540

Query: 542 LPESVQEIGRSAPFRQNGALHLMFIGNKVKITIGEMAFLSNKLSEVNLSQKQLKTIEVQAF 601
 LPESVQEIGRSAPFRQNGA +L+F+G+KVKT+GEMAFLSN+LE ++LSEKQL I VQAF
 Sbjct: 541 LPESVQEIGRSAPFRQNGANNLIFMGSKVKTIGEMAFLSNLEHLDLSEKQLTEIPVQAF 600

Query: 602 SDNALSEVVLPPNLQTIREEAFKRNHLKEVKSSTLSQITFNAFDQNDGDKRFGKKVVVR 661
 SDNAL EV+LP +L+TIREEAFK+NLK+++ +S LS I FNA D NDGD++F KVVV+
 Sbjct: 601 SDNALKEVLLPASLKTIREEAFKRNHLKQLEVASALSHIAPNALDDNDGDEQFDNKVVVK 660

Query: 662 THNNSHMLADGERFIIDPKLSSTMVDLEKVLKIEGLDYSLRQTQTQFRMTTAGKA 721
 TH+NS+ LADGE FI+DPDKLSST+VDLEK+LK+IEGLDYSLRQTQTQFR+MTTAGKA
 Sbjct: 661 THNSYALADGEHFIVDPDKLSSTIVDLEKILKIEGLDYSLRQTQTQFRDMMTAGKA 720

Query: 722 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGHLLERSINKAVL 781
 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNG LLERSINKAVL
 Sbjct: 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780

Query: 782 AYNNSAIKKANVKRLEKELDLTDLVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 841
 AYNNSAIKKANVKRLEKELDLT LVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY
 Sbjct: 781 AYNNSAIKKANVKRLEKELDLTGLVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 840

Query: 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYHTLAVATLADYEGLYIKDILN 901
 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYH LAVATLADYEG L IK ILN
 Sbjct: 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVEDNEGYHALAVATLADYEGLDIKTILN 900

Query: 902 SSLDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLNEVPNRYKKQVEKNL 961
 S L ++ +IRQ+P A YHR GIFQAI+NAAAEA++LLPK
 Sbjct: 901 SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEABQLLPK----- 939

Query: 962 KPV DYKTPIFNKALPNEKVDGDRAAKGHNINAETNNSVAVTPIRSEQQLHKSQSDVNLPQ 1021
 ++++ + N++ ++S + ++ + LP+
 Sbjct: 940 -----PGTHSEKSSSESANSKDRG-----LQSNPKTNRGRHSAILPR 977

Query: 1022 TSSKNFIYEILGYVSLCLLFLVTAGKKGK 1051
 T SK +F+Y ILGY S+ LL L+TA KK K
 Sbjct: 978 TSGKGSFVYGILGYTSVALLSLITAIKKKK 1007

SEQ ID 800 (GBS97) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 12; MW 113.4kDa).

-318-

GBS97-His was purified as shown in Figure 193, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 254

- 5 A DNA sequence (GBSx0269) was identified in *S. agalactiae* <SEQ ID 803> which encodes the amino acid sequence <SEQ ID 804>. This protein is predicted to be ribonucleoside-diphosphate reductase alpha chain (nrdE). Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4274 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB96160 GB:AE000050 ribonucleoside-diphosphate reductase alpha
chain-MPN324 (new), 513 (Himmelreich et al., 1996)
[Mycoplasma pneumoniae]

20

Identities = 476/725 (65%), Positives = 586/725 (80%), Gaps = 20/725 (2%)

Query: 2 TQSD--AYLSLNAKTRFRDRITGNYHFTSDKEAVEQYMIHVPEPTMVFTSLIEKLDYLV 59
TQ D +Y+SLNA T+ F D AVE Y+ EHV+P T VF S E+LD+LV

25

Sbjct: 12 TQEDLESYISLNAYTKVYG-----DFKMDLHAVEAYIQEHVKPKTKVPHSTKERLDLFLVK 66

Query: 60 NNYVESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAVALKTEDNRYYLEHYED 119
N+YY+ +++ Y+ E +I AYA +F + NFMGA KFYNAVALKT D ++YLE+YED

Sbjct: 67 NDYYDENIINMYSFEQFEETRKAYAYRFRYANFMGAFKFYNAVALKTFDQKYLENYED 126

30

Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRGEYISCYLLRIEDNME 179
RVVMN LFLA G+ A L+ ++ NRFQPATPTFLNAG+K+RGE++SCYLLRIEDNME

Sbjct: 127 RVVMNVFLANGNYNKALKLLKQIITNRFQPATPTFLNAGKKRGEFVSCYLLRIEDNME 186

35

Query: 180 SISRAISTSLQSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLEDSFSYANQL 239
SI RAI+T+LQLSKR GGVAL LTN+RE GAPIK I+NQ++GI+P+MKLEDSFSYANQL

Sbjct: 187 SIGRAITTTTLQLSKRDGGVALLLTNIRESGAPIKKIENQSSGIIPIMKLEDSFSYANQL 246

40

Query: 240 GQRQGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLSLGLVIPDITFELAKANKDMAL 299
GQRQGAGAVYLHAHH+V+ FLDTKRENADEKIRIKSLSLGLVIPDITF LAK N++MAL

Sbjct: 247 GQRQGAGAVYLHAHHPDVMQFLDTKRENADEKIRIKSLSLGLVIPDITFTLAKNNEEMAL 306

45

Query: 300 FSPYDIERVYKPMDSISITEEYETLLANADIRKTFISARKLFQTIAELHFESGYPYILF 359
FSPYD+ YGKP+SDIS+TE Y LLAN I+KTFI+ARK FQT+AEELHFESGYPYILF

Sbjct: 307 FSPYDVYEEYKPLSDISVTEMYELLANQRIKKTFINARKFFQTVAELHFESGYPYILF 366

Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHDVCCNLGSINIARAM 419
+DTVN +N H RIVMSNLCSEI Q +T S+F DL F KVG+D+ CNLGS+NIA+AM

Sbjct: 367 DDTVNRNRNAH--PNRIVMSNLCSEIVQFSTPSEPHDLAFKKVGNDISCNLGSNLIAKAM 424

50

Query: 420 DQAADFEKLIANSIRALDRVSRTSDLDSAPSIKKGNAANHAVGLGAMNLHGFLATNHIYY 479
+ +F +L+ +I +LD VSR S+L++APSI+KGN+ NHA+GLGAMNLHGFLATN IYY

Sbjct: 425 ESGPEFSELVKLAIESLDLVSRSVNSLETAPSIQKCNSENHALGLGAMNLHGFLATNQIYY 484

55

Query: 480 DSQEADFTDCFFYAMAYYAFKASNHLAKEKGTFEFGFSESSYADGSYFYQY--TEQNF-E 536
+S EADFT+ FFY +AY+AFKAS+ LA EKG F+ F + +ADGSYF +Y E +F

Sbjct: 485 NSPEADFTNIFFFYTVAYHAFKASSELAKKGFKNFENTKFDGYSYFDKYIKVEPDFWT 544

Query: 537 PKTORVKNLAEYGLTLPSQEDWRKLVSQSIKEIGLANAHLAVAPTGSISYLSSTCTPSLQ 596
PKT+RVK L +Y + +P++E+W++L +I++ GLAN+HLLA+APTGSISYLSSTCTPSLQ

60

Sbjct: 545 PKTERVKALFQKYQVEIPTRENWKEALNIQKNGLANSHLLAIAPTGSISYLSSTCTPSLQ 604

-319-

Query: 597 PVVSPVEVRKEGALGRVYVPAYKIDADNYVYKKGAYEVGSEAIINIAAAQKHIDQAIS 656
 PVVSPVEVRKEG LGR+YVPAY+++ D+Y +YK GAYE+G E IINIAAAQ+H+DQAIS
 Sbjct: 605 PVVSPVEVRKEGRLGRIYVPAYQLNKDSYFFYKDGAYELGPEIINIAAAQQHVDQAIS 664

Query: 657 LTLFMTDQATTRDLNKAYIQAFKQKASIIYVVRVQDILEGSESYDDMLDDFTSSDLEDC 716
 LTLFMTD+ATTRDLNKAYI AFK+ C+SIIYVVRVQ++LE SE + + ++ C
 Sbjct: 665 LTLFMTDKATTRDLNKAYIYAFKKGCSIIYVVRVQEVLEDSH-----TIQMQQC 716

Query: 717 QSCMI 721
 ++C+I
 Sbjct: 717 EACVI 721

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 805> which encodes the amino acid
 sequence <SEQ ID 806>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides forsythus]
 Identities = 124/451 (27%), Positives = 202/451 (44%), Gaps = 65/451 (14%)

Query: 221 FNSYLLKKVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFAHLA-LKQIDLPDN 279

Sbjct: 87 F+ L VT+P IG AF + + +P S+ TI ++AF + LK I LP++
 FSDCALTSVTLPLNSLTAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPLNS 146

Query: 280 LKAIGELAFFDNQITGKLSLRLMLRAERA-FKSNHIKTIEFRGNSLKVIGEASFQD-N 337
 L AIG+ A +++P + + E A F + + +I F NSL IGE++F

Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFKGCGLTSITF-PNSLTAIGESAFYGC 205

Query: 338 DLSQLMLPDGLEKIESEAFTEGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQ 397
 L+ + LPD L I AF G G KS P+ L T +S +

Sbjct: 206 ALTSITLPDALTTIGESAFKGCGLSITITFPNSLTTIG-----ESAFY 248

Query: 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGKGLQKVRNKNLEIPKHNGVTITEIGDNAPRN 457
 + + + T +++ G S GL K++ P ++T IG++AF N

Sbjct: 249 DCGALTSITLPDALTTIGRSAPYGCGLSITITFPN-----SLTTIGESAFYN 295

Query: 458 VDFQNKTLRKYDLEEVKLPSTIRKIGAFAPQS-NNLKSFEASDDLEEIKEGAFMNNRIET 516
 L + +P+++ IG AF + LKS D L I+E AF N + T

Sbjct: 296 CG-----SLTSITIPNSVTTIGRSAPYGCGLKSITLPDGLTTIEERAFYNGVLT 346

Query: 517 -LELKDKLVTIGDAAFH-INHIYAIVLPESVQEIGRSAPRQNGANNLI FMGSKVKTILGEM 574
 + + + + TIG++AF+ + + +I LP+ + I AF GA I + + V T+GE

Sbjct: 347 SITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAFYNGALTSITIPNSVSTIGES 406

Query: 575 AFLS-NRLEHDLSEKQLTEIPVQAFSDNALKEVLL--PASLKTIREEAFFKNHLKQLE 631
 AF L+ + + + +I F + L + L PA KT+ E K+ K+

Sbjct: 407 AFYGCGLKDVTVAVNDTPI-DIQRDVFRELTLSGIRLHVPAGKKTVE--AKDVWKE-- 460

Query: 632 VASALSHIAFNALDDND-GDEQFDNKVVVKT 661
 FN ++D+D G Q++ KT

Sbjct: 461 -----FNIVEDDDFGGLQWNYDAATKT 482

An alignment of the GAS and GBS proteins is shown below:

Identities = 534/726 (73%), Positives = 614/726 (84%), Gaps = 5/726 (0%)

Query: 1 MTQSDA-YLSLNAKTRFRDRGTGNYHFTSDKEAVEQYMIHVEPNTMVFTSLIEKLDYLV 59

-320-

M+Q++A YLSLNA TRF+ G+YHF SDKEAV +Y+ EHV PN M F SL +KL YL++
 Sbjct: 1 MSQTNASYLSLNA/TRFKKPDGSHYFSDSDKEAVRRYLEEHVSPNQMAFNSLEDKLAYLIN 60
 Query: 60 NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAVALKTEDNRYLLEHYED 119
 YYE + Y + I + F +AY + + FLN MGA+KFY +YALKT D + YLE +ED
 Sbjct: 61 EGYEQAIIFDAYPNDLIKEAFHYAYQGGYRFLNLMGAMKFYQSYALKTLGDKQYLETFFED 120
 Query: 120 RVMNALFLAAGDEKAAAYDLVDDMLANRFQPATPTFLNAGKKRRGEYISCYLLRIEDNME 179
 R VMNALFLA GD+ +D++D +L RFQPATPTFLNAGKKRRGEYISCYLLR+EDNME
 Sbjct: 121 RAVMNALFLADGDQTFVFDVIDAILHRRFQPATPTFLNAGKKRRGEYISCYLLRVEDNME 180
 Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLEDSFSYANQL 239
 SISRAISTSLQLSKRGGGVALCLTNLRE GAPIKGI+NQATGIVPVMKLEDSFSYANQL
 Sbjct: 181 SISRAISTSLQLSKRGGGVALCLTNLREIGAPIKGIENQATGIVPVMKLEDSFSYANQL 240
 Query: 240 GQRQAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLGLVIPDITFELAKANKDMAL 299
 GQRQAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSL+LGLVIPDITF+LAK NKDMAL
 Sbjct: 241 GQRQAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLALGLVIPDITFQLAKENKDMAL 300
 Query: 300 FSPYDIERVYGKPMDSISITEEYETLLANADIRKTFISARKLFQTTAELHFESGYPYILF 359
 FSPYDI+R YGK MSDISITEEY+ LLAN I+KT+ISARK FQ IAEELHFESGYPY+LF
 Sbjct: 301 FSPYDIKRAYGKMDSDISITEEYDKLLANPAIKKTYISARKFFQLIAELHFESGYPYILF 360
 Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHDVCCNLGSINIAM 419
 +DTVN +NPH K+GRIVMSNLCSEIAQV+T S F EDL+F +G D+CCNLGSINIA+AM
 Sbjct: 361 DDTVNKRNPFAKGRIVMSNLCSEIAQVSTPSTFKEDLSFETIGEDICCNLGSINIAQAM 420
 Query: 420 DQAADEFKLIANSIRALDRVSRTSDLDAPSISIKKGNANHAVGLGAMNLHGFLATNHIYY 479
 A FE+LI SIRALDRVSR SDL+ APS++ GNAANHAVGLGAMNLHGFLATNHIYY
 Sbjct: 421 ADAPHFEQLITTSIRALDRVSRVSDLNCAPSVETGNAANHAVGLGAMNLHGFLATNHIYY 480
 Query: 480 DSQEAIIDFTDCFFYAMAYYAFKASNHLAKEKGTFFEGFSESSYADGSYFYQTEQNFEPKT 539
 D++EA+DFTD FF+AMAYYAFKAS LAKEKG F GFS S+Y+DG+YF +Y +++ +P+T
 Sbjct: 481 DTKEAVDFTDLFFHAMAYYAFKASCQLAKEKGAFAGFSLSTYSDGTYFAKYLQEDAKPQT 540
 Query: 540 QRVKNLLAEYGLTLPQEDWRKLVQSIKEIGLANAHLAVAPTGSISYLSSTPSSLQPVV 599
 +V LL +YG TLP+ DW+ LV IK+ GLANAHLAVAPTGSISYLSSTPSSLQPVV
 Sbjct: 541 AKVATLLQDYGFTLPTVADWQALVADIKQFGLANAHLAVAPTGSISYLSSTPSSLQPVV 600
 Query: 600 SPVEVRKEGALGRVYVPAYKIDADNYVYKKGAYEVGSEAIINIAAAQKHIDQAISLTL 659
 +PVEVRKEG+LGR+YVPAY+ID NY YY++GAYEVG +AII++ AAAQKH+DQAISLTL
 Sbjct: 601 APVEVRKEGSLGRIYVPAYQIDQANYAYYERGAYEVGPKAIIIDVVAQAQKHVDQAISLTL 660
 Query: 660 FMTDQATTRDLNKAYIQAFKQKASIYYVRVRQDILEGSESYDD---MLDDFTSSDLED 715
 FMTDQATTRDLN++YIQAFKQ CASIYYVRVRQD+L GSE YD+ + +
 Sbjct: 661 FMTDQATTRDLNRSYIQAFKQNCASIYYVRVRQDVLGSEQYDEDSLVTPAGASDETTTE 720
 Query: 716 CQSCMI 721
 CQSCMI
 Sbjct: 721 CQSCMI 726

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 255

A DNA sequence (GBSx0270) was identified in *S.agalactiae* <SEQ ID 807> which encodes the amino acid sequence <SEQ ID 808>. This protein is predicted to be nrdI protein (nrdI). Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2952(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-321-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAC71451 GB:U39702 nrdI protein (nrdI) [Mycoplasma genitalium]
 Identities = 77/127 (60%), Positives = 104/127 (81%), Gaps = 1/127 (0%)

Query: 7 VVYFSSKSNNTHRFVQKLACSNQRIPSD-GSSILVTEDYILIVPTYAGGGDDTKGAVPKQ 65
 +VYFSS SNNTHRF++KL ++RIP D SI V+ +Y+LI PTY+GGG+ +GAVPKQ
 10 Sbjct: 22 IVYFSSISNNTHRFIEKLGFOHKRIPVDITQSITVSNEYVLICPTYSGGGNQVEGAVPKQ 81

Query: 66 VVQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVP LLHQFELLGTQEDVTRV 125
 V+QFLN + NRE C+GVI+SGNTNFGDT+ +AG +I++KLNVP LL+QFELLGT+ DV +
 Sbjct: 82 VIQFLNNKHNRCLCRGVIASGNTNFGDTFCLAGTVISKLNVP LLVQFELLGTKNDVEQT 141

15 Query: 126 KELLQCF 132
 ++++ F
 Sbjct: 142 QKIINF 148

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 809> which encodes the amino acid
 sequence <SEQ ID 810>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0089(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 84/125 (67%), Positives = 100/125 (79%)

Query: 7 VVYFSSKSNNTHRFVQKLACSNQRIPSDGSSILVTEDYILIVPTYAGGGDDTKGAVPKQV 66
 +VYFSSKSNNTHRFVQKL QRIP D + V+ Y+LIVPTYA GG D KGAV KQV
 35 Sbjct: 6 IVYFSSKSNNTHRFVQKLGLPAQRIPVDNRPLEVSTHYLLIVPTYAAGGSDAKGAVSKQV 65

Query: 67 VQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVP LLHQFELLGTQEDVTRVK 126
 ++FLN NR+HC+GVISSGNTNFGDT+A+AGPII++KL VPLLHQFELLGT DV +V+
 40 Sbjct: 66 IRFLNPNNRKHCKGVISSGNTNFGDTFALAGPIISQKLQVPLLHQFELLGTATDVKKVQ 125

Query: 127 ELLCQ 131
 + +
 Sbjct: 126 AIFAR 130

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 256

A DNA sequence (GBSx0271) was identified in *S.agalactiae* <SEQ ID 811> which encodes the amino acid
 sequence <SEQ ID 812>. This protein is predicted to be ribonucleoside-diphosphate reductase beta chain
 (nrdF). Analysis of this protein sequence reveals the following:

50 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3889(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-322-

>GP:AAB96162 GB:AE000050 ribonucleoside-diphosphate reductase beta
chain [Mycoplasma pneumoniae]
Identities = 261/335 (77%), Positives = 301/335 (88%)

5 Query: 2 QSYDRSQSPLDYALSEKAFPMRSVNWKLNDKDELVWNRVTQNFWLPEKIPVSNLNS 61
+ Y+ S SPL+YA + +RSVWNW ++D+KDLEVWNR+TQNFWLPEKIPVSN+ S
Sbjct: 5 KKYFLESVSPLEYAQKFKQGNLRSVNWNLVDDEKDELVWNRITQNFWLPEKIPVSNLIPS 64

10 Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121
W+ L +WQ LIT+TFTGLTLLD++QAT+GDI QI ++ TDHEQVIYANFAFMV +HARS
Sbjct: 65 WKQLSKEWQDLITKTFTGLTLLDTIQATIGDIKQIDYALTDHEQVIYANFAFMVGVHARS 124

15 Query: 122 YGTIFSTLCTSQIEEAHEWVVDTESLQARSRLIPFYTGDDPLKSKVAAAMMPGFLLYG 181
YGTIFSTLCT+QI EAHEWVV TESLQ R++ LIP+YTG DPLKSKVAAA+MPGFLLYG
Sbjct: 125 YGTIFSTLCTSEQITEAHEWVVKTESLQKRAKALIPYYTGKDPLKSKVAAALMPGFLLYG 184

20 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYSGYKQKQKVAKLSVEKQAEKMTFVFD 241
GFYLPFYLS+R +LPNTSDIIRLILRDKVIHNYSGYK+Q+KV K+S EKQAEK FVFD
Sbjct: 185 GFYLPFYLSRRKQLPNTSDIIRLILRDKVIHNYSGYKQKQKVAKLSVEKQAEKMTFVFD 244

25 Query: 242 LLYQLIDLEKAYLYELYDGFDAEDAIRFSIYNAGKFLQNLGYDSPFTEETRISPEVFA 301
L+Y+LI+LEKAYL ELY+GF + EDAI+FSIYNAGKFLQNLGYDSPFTEETRI PE+FA
Sbjct: 245 LMYELIELEKAYLKELYEGFIVEDAIKFSIYNAGKFLQNLGYDSPFTEETRIKPEIFA 304

30 Query: 302 QLSARADENHDFSGNGSSYIMGITEETLDEDF 336
QLSARADENHDFSGNGSSY+MGI+EET D+DW+F
Sbjct: 305 QLSARADENHDFSGNGSSYVMGISEETEDKDWDF 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 813> which encodes the amino acid
sequence <SEQ ID 814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3779(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 292/335 (87%), Positives = 318/335 (94%)

Query: 2 QSYDRSQSPLDYALSEKAFPMRSVNWKLNDKDELVWNRVTQNFWLPEKIPVSNLNS 61
Q YY+RSQSP++YALSE +RS+WNW LNDDKDELVWNRVTQNFWLPEK+PVSNLNS
Sbjct: 3 QHYERSQSPIEYALSETQQLRSINWNYLNDDKDELVWNRVTQNFWLPEKVPVSNLNS 62

45 Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121
WR+L DWQQLITRT+TGLTLLD+VQATVGD+AQI+HSQTDHEQVIY NFAFMV IHARS
Sbjct: 63 WRSLGEDWQQLITRTYTGLTLLDTVQATVGDVAQIQHSQTDHEQVIYTNFAFMVGIHARS 122

50 Query: 122 YGTIFSTLCTSQIEEAHEWVVDTESLQARSRLIPFYTGDDPLKSKVAAAMMPGFLLYG 181
YGTIFSTLC+S+QIEEAHEWVV T+SLQ R+R+LIP+YTGDDPLKSKVAAAMMPGFLLYG
Sbjct: 123 YGTIFSTLCSSEQIEEAHEWVVSTQSLQDRARVLIPYYTGDDPLKSKVAAAMMPGFLLYG 182

55 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYSGYKQKQKVAKLSVEKQAEKMTFVFD 241
GFYLPFYLSARGK+PNTSDIIRLILRDKVIHNYSGYKQKQKVA+LS EKQAEK FVFD
Sbjct: 183 GFYLPFYLSARGKMPNTSDIIRLILRDKVIHNYSGYKQKQKVARLSPEKQAEKAFVFD 242

60 Query: 242 LLYQLIDLEKAYLYELYDGFDAEDAIRFSIYNAGKFLQNLGYDSPFTEETRISPEVFA 301
LLY+LIDLEKAYL ELY GFDLAEDAIRFS+YNAGKFLQNLGY+SPFT+EETR+SPEVFA
Sbjct: 243 LLYELIDLEKAYLRELYAGFDLAEDAIRFSLYNAGKFLQNLGYESPFTEETRVSPEVFA 302

65 Query: 302 QLSARADENHDFSGNGSSYIMGITEETLDEDF 336
QLSARADENHDFSGNGSSY+MGITEET D+DWF
Sbjct: 303 QLSARADENHDFSGNGSSYVMGITEETDDDF 337

-323-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 257

A DNA sequence (GBSx0272) was identified in *S.agalactiae* <SEQ ID 815> which encodes the amino acid sequence <SEQ ID 816>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1741(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9583> which encodes amino acid sequence <SEQ ID 9584> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
Identities = 104/309 (33%), Positives = 173/309 (55%), Gaps = 21/309 (6%)

Query: 11 QINICLATYNGQKYLRLQQLDSIIQGGYTDWICLIRDDGSTDDTVVAIIKEYVNRDSRFIFI 70
++NI ++TYNGQ+++ QQ+ SI +Q + +W LIRDDGS+D T II ++ D+R FI
Sbjct: 2 KVNILMSTYNGQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPTIADFAKSDARIRFI 61

Query: 71 NSNDRKLGSHRSFYELVNYKKADFYVFSQDDVWKENRLERYLEEAEKFNQELPLLVS 130
N++ G ++FY L+ Y+KAD+Y FSDQDDVW +LE L EK N ++PL+VY+
Sbjct: 62 NADKRENFVGIKNFYTLKYEKADYFFSDQDDVWLPQKLELTLASVEKENNQIPLMVYT 121

Query: 131 NWTSVDEKLTVL-----KEHNPATVIOEQIAFNQINGMVMIMNHELAKLWE--YRQIG 181
+ T VD L VL + H+ T + E++ N + G +M+NH LAK W+ Y +
Sbjct: 122 DLTVVDRDLQVLHDSMIKTQSHHANTSLEELTENTVTGGTMMVNHCLAKQWKQCYDDLI 181

Query: 182 AHDYVGTLAYAVGNVAYISDSTVLWRRQ----VGAES----LNNGRQYG-VATFWQMI 232
HD Y+ LA ++G + Y+ ++T L+R+ +GA + L N+ R + V +W ++
Sbjct: 182 MHDWYLAALLAASLGKLIYLDETTELYRQESNVLGARTWSKRLKNWLRPHRLVKKYWWLV 241

Query: 233 NTSFDRASLIFAQVSDKMSLERKLFSSRFIELKNANLMRRIYLLSKLRLRKSLSKETVAM 292
+S +AS + + + K ++ L + + + RI L + + T
Sbjct: 242 TSSQQQASHL---LELDLPAANKAIRAYVTLLDQSFLNRIKWLKQYGFKNRAFHTFVF 298

Query: 293 TILLITGYG 301
L++T +G
Sbjct: 299 KTLIITKFG 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 258

A DNA sequence (GBSx0273) was identified in *S.agalactiae* <SEQ ID 819> which encodes the amino acid sequence <SEQ ID 820>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.19 Transmembrane 1213 -1229 (1211 -1230)

----- Final Results -----
bacterial membrane --- Certainty=0.2678(Affirmative) < succ>

-324-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 A related GBS nucleic acid sequence <SEQ ID 9581> which encodes amino acid sequence <SEQ ID 9582> was also identified.

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8525> and protein <SEQ ID 8526> were also identified. Analysis of this protein sequence reveals the following:

10 Lipop: Possible site: -1 Crend: 7
 SRCLG: 0
 McG: Length of UR: 3
 Peak Value of UR: 2.28
 Net Charge of CR: 4
 15 McG: Discrim Score: 1.29
 GvH: Signal Score (-7.5): 2.84
 Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 31
 20 ALOM program count: 0 value: 1.16 threshold: 0.0
 PERIPHERAL Likelihood = 1.16 344
 modified ALOM score: -0.73
 *** Reasoning Step: 3
 25 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
 30 LPXTG motif: 1197-1201

SEQ ID 8526 (GBS147) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 4; MW 132kDa).

35 The GBS147-His fusion product was purified (Figure 200, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 286), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 259

40 A DNA sequence (GBSx0274) was identified in *S.agalactiae* <SEQ ID 823> which encodes the amino acid sequence <SEQ ID 824>. This protein is predicted to be Acetyltransferase (GNAT) family. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2781 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG03505 GB:AE004449 conserved hypothetical protein [Pseudomonas aeruginosa]

-325-

Identities = 66/143 (46%), Positives = 94/143 (65%), Gaps = 5/143 (3%)

Query: 2 WNVKTFDNLTHLTFQIYKLRVSFVVEQDCPYQEVDDEDLI--CLHGMNWDGQLAAYY 59
 W K +LT EL+ + +LR VFWVEQ CPYQEVD DL+ H M W DGQL AY
 5 Sbjct: 5 WTCKHHADLTLELYALLQLRTEVFVVEQKCPYQEVDGLDLVGDTHHLMAWRDGQLLAYL 64

Query: 60 RLIP---EDDKVHLGRVIVNPDFRKKGLGNQLVEYAIKFSEANYPNKPIYAQAQAYLQDF 116
 RL+ + +V +GRV+ + R +GLG+QL+E A++ +E + + P+Y AQA+LQ +
 10 Sbjct: 65 RLLDPVRHEGQVVIGRVVSSAARGQGLGHQLMERALQAAERLWLDTPVYLSAQAHQAY 124

Query: 117 YQSFQFQPVSDIYLEDNIPHLDM 139
 Y +GF V+++YLED+IPH+ M
 Sbjct: 125 YGRYGFVAVTEVYLEDDIPHGM 147

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 260

20 A DNA sequence (GBSx0275) was identified in *S.agalactiae* <SEQ ID 825> which encodes the amino acid sequence <SEQ ID 826>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2010(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 261

35 A DNA sequence (GBSx0276) was identified in *S.agalactiae* <SEQ ID 827> which encodes the amino acid sequence <SEQ ID 828>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAB12631 GB:Z99108 similar to RNA methyltransferase [Bacillus subtilis]
 Identities = 217/448 (48%), Positives = 298/448 (66%), Gaps = 4/448 (0%)

Query: 7 QRIPLKIKRMGINGEGIGFYKKTILIFVPGALKGEEVFCQISSVRRNF AEAKLLKINKKSK 66
 Q PL IKR+GINGEG+G++KK ++FVPGAL GEEV Q + V+ F+E ++ KI K S+
 50 Sbjct: 16 QTFPLTKRLGINGEGVGYFKKKVVFVPGALPGEEVVQATKVQPKFSEGRICKIRKASE 75

Query: 67 NRVEPPCSIYKECGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKTI GMSSEP 126
 +RV PPC +Y++CGGCQ+ HL Y +QL K D++ Q+L + EN EI++TIGM P

-326-

Sbjct: 76 HRVAPPCPVYEQCGGQQLQHLAYSQQLREKRDIVIQSLERHTKFKVENMEIKETIGMDNP 135

Query: 127 EHYRAKLQFQV-RSFGGNVKAAGLYAQGTHRLIDIKDCLVQDSLQEMINRVAELLGKYKL 185
 +YR K QFQ+ RS G++ AGLY +H ++ IKDC+VQ T + V +L + +

5 Sbjct: 136 WNYRKSQFQIGRSQSGSIIAGLYGLDSDHIVPIKDCTIVQHPATNKTGIVRRILEDFNV 195

Query: 186 PIYNERKIAG-VRTVMIRRAQASGEVQLIFITTSKRL--DFDDVVIELVREFPELKTVAVN 242
 +YNERK G VRT++ R +GEVQ++ +T+K +++V + + PE+K++ N

10 Sbjct: 196 SVYNERKRGDVRTIVRVGFETGEVQVVLVTAKETLPHKEEIVKAIQKRLPEVKSIIQN 255

Query: 243 INASKTSDIYGQITEVIWQESINEEVLGYGFSLSPRAFYQLNPKQTQILYSEAVKALDV 302
 +N +KTS I+G+ T+ + G+ I E + D F LS RAF+QLNP+QT LY E KA +

Sbjct: 256 VNGAKTSVIFGEKTKQLAGKTVIQEVLGDVVSFELSARAFFQLNPEQTVKLYDEVKAAQL 315

15 Query: 303 KEDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIPEAIQDAKENALYMGFTTNTHYEAGKAE 362
 + ++DAYCGVGTIG+ A K VRGMD+I E+I DAK+NA G N Y G AE

Sbjct: 316 TGKEKVVDDAYCGVGTIGMWVADGAKEVRGMDVIKESIDDAKNAKKGHMANATYVTGTAE 375

Query: 363 DIIPRWYSEGFRANALIVDPPTGLDDKLLNTILKMPPEKMYVSCNTSTLARDLVTLTK 422
 +P+W EGFR + +IVDPPRTG D L+TI K+ P++ VYVSCN STLA+DL TL+K

20 Sbjct: 376 HWLPKWTKEGFRPDVIVDPPTGCDSTFLDTIKVKPKRFVYVSCNPSTLAKDLQTLTK 435

Query: 423 VYHVHYIQSVDMFPHTARTEAVVKLQK 450
 Y V YIQ VDMFP TA EAV +L K

25 Sbjct: 436 DYRVDIQFVDMFPQTAHVEAVARLVLK 463

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 829> which encodes the amino acid sequence <SEQ ID 830>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2980(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

An alignment of the GAS and GBS proteins is shown below:

Identities = 327/450 (72%), Positives = 397/450 (87%)

40 Query: 1 MNVVLKQRIPLKIKRMGINGEGIGFYKKTILFVPGALKGEEVFCQISSVRRNFAEAKLLK 60
 M V +KQ+IPLKIKRMGINGEGIGFY+KTL+FVPGALKGE++FCQI++V+RNFAEAKLL
 Sbjct: 1 MVVVKQKQKIPLKIKRMGINGEGIGFYQKTLVFPVPGALKGEDIFCQITAVKRNFAEAKLLT 60

45 Query: 61 INKSKNRVEPPCSIYKECGGQIMHLYQDKLEFKTDVIRQALMKFKPEGYENYEIRKT 120
 +NK SKNRV+P CS+Y+ CGGQIMHL Y KQL+FK DVIQAL KFKP GYE +EIR T
 Sbjct: 61 VNKASKNRVKPACSVYETCGGQIMHLYPKQLDFKDDVIRQALKKFKPTGYEQFEIRPT 120

Query: 121 IGMSEPEHYRAKLQFQVRSFGGNVKAAGLYAQGTHRLIDIKDCLVQDSLQEMINRVAELL 180
 +GM +P+HYRAKLQFQ+RSFGG VKAGL++QG+HRL+ I +CLVQD LTQ++IN++ +L+

50 Sbjct: 121 LGMKKPDHYRAKLQFQLRSFGGTVKAGLFSQGSRLVPIDNCLVQDQLTQDIINKITQLV 180

Query: 181 GKYKLPIYNERKIAGVRTVMIRRAQASGEVQLIFITTSKRLDFDDVVIELVREFPELKTVA 240
 KYKLPIYNERKIAG+RT+M+R+AQAS +VQ+I ++SK + + + EL + FP++KTVA

55 Sbjct: 181 DKYKLPIYNERKIAGIRTIMVRKAQASDQVQIIVSSKEVRLANFIGELTKAFPQVKTVA 240

Query: 241 VNINASKTSDIYGQITEVIWQESINEEVLGYGFSLSPRAFYQLNPKQTQILYSEAVKAL 300
 +N N SK+S+IYG TE++WGQE+I+EEVLGYG+LSPRAFYQLNP+QT++LY E VKAL

Sbjct: 241 LNSNRKSSSEIYGDETEILWGQEAIIHEEVLGYGFALSPPRAFYQLNPQQTEVLYGEVVKAL 300

60 Query: 301 DVKEDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIPEAIQDAKENALYMGFTTNTHYEAGK 360
 DV D +IDAYCGVG+IG AFAGKVKSVRGMDIPEAI+DA++NA MGF N +YEAGK

Sbjct: 301 DVGSKDHIIIDAYCGVGSIGFAFAGKVKSVRGMDIPEAIEDAQKNAKMGFDNAYYEAGK 360

65 Query: 361 AEDIIIPRWYSEGFRANALIVDPPTGLDDKLLNTILKMPPEKMYVSCNTSTLARDLVTL 420
 AEDII +WY +G+RA+A+IVDPPRTGLDDKLL TIL P++MVVSCNTSTLARDLV L

-327-

Sbjct: 361 AEDIISKWKYQGYRADAVIVDPPTGLDDKLLKLTILHYQPKQMVYVSCNTSTLARDLVQL 420

Query: 421 TKVYHVHYIQSVDMFPHTARTEAVVKLQRK 450
TKVY VHYIQSVDMFPHTARTEAVVKLQ++

5 Sbjct: 421 TKVYDVHYIQSVDMFPHTARTEAVVKLQKR 450

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 262

- 10 A DNA sequence (GBSx0277) was identified in *S.agalactiae* <SEQ ID 831> which encodes the amino acid sequence <SEQ ID 832>. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3505(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04643 GB:AP001510 unknown conserved protein in B. subtilis
[Bacillus halodurans]
Identities = 74/263 (28%), Positives = 141/263 (53%), Gaps = 9/263 (3%)

25 Query: 3 ITKIEKKKR---LYTLEL-DNTENLY---ITEDTIVHFMLSKGMIINAEKLENIKKFAQL 55
IT+IE +KR Y + + N +++Y + E ++ L KG+ I+AE+++ I ++
Sbjct: 4 ITRIEVQKRNNERYNIFIHQNGQDVYAFSDVEQVLIKQGLRKGLDIDAEQMKQILYEDEV 63

30 Query: 56 SYGKNLGLYYISFKQRTKEVIKYLQOHDIDSKIIPQIIDNLKSENWINDKNYVQSFIQQ 115
NL L+Y+S++ R+ EV YL++ D + II ++ L + ++D + ++FIQ
Sbjct: 64 QKTFNLALHYLSYMRMSVHEVRTYLKKKDREEPIIEHVLHRLTEQRLDDHAFARAFIQ 123

35 Query: 116 NLNTGDKGPGYVIKQKLLQKGIKSIIESELQAINFQDLASKISQKLYKKYQNKLPKAL- 174
T KGP +KQ+L +KG+ K IE L ++++ ++ L K+ +L
Sbjct: 124 KRATTSKGPLKQELAEGVSEKTIEGALTTPSYEEQVEQVKAWEKQKGRFTKGSSLA 183

40 Query: 175 -KDKLMQSLTTKGFQYIVHTVIQNLEIEKDQELEEDLIYKELDKQYQKLSKKHDQYELK 233
K KL + L KG+ ++ ++ I++++E E + + +K +K + K +EL+
Sbjct: 184 WKQKLSRQLLAKGYTSPVIEEAFADVPKQEEEEWEALKAFGEKAMRKYAGKKTGWELQ 243

Query: 234 QRINALMRKGYQYEDIKSALRE 256
Q++ AL RKG+ E I+ L +
Sbjct: 244 QKVKQALYRKGFSGLEMIERYLND 266

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 833> which encodes the amino acid sequence <SEQ ID 834>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2388(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 55 An alignment of the GAS and GBS proteins is shown below:

Identities = 146/258 (56%), Positives = 190/258 (73%)

Query: 1 MKITKIEKKKRLYLTLELDNTENLYITETDITIVHFMLSKGMIINAEKLENIKKFAQLSYGKN 60
MKITKIEKKKRLY +ELDN E+LY+TETDITIV FMLSK +++ ++LE++K FAQLSYGKN

-328-

Sbjct: 1 MKITKIEKKRLYLIELDNDESLVYVTEDTIVRFMLSXDKVLDNDQLEDMKHFAQLSYGKN 60

Query: 61 LGLYYISFKORTEKEVIKYLQOHDIDSKIIPQIIDNLKSENWINDKNVQSFIQQNLNTG 120
L LY++SF+QR+ K+V YL++H+I+ II II L+ E WI+D ++I+QN G

5 Sbjct: 61 LALYFLSFQQRSNKQVADYLRKHEIEEHIIADIITQLQEEQWIDDTKLADTYIRQNQLNG 120

Query: 121 DKGPYVIKQKLLQKGIKSIKIESLQAINFQDLASKISQKLYKKYQNKLPKALKDKLMQ 180
DKGP V+KQKLLQKGI S I+ L +F LA K+SQKL+ KYQ KLP KALKDK+ Q

10 Sbjct: 121 DKGPQVLKQKLLQKGIASHDIDPILSQITDFSQLAQKVSQKLFDKYQEKLPKALKDKKITQ 180

Query: 181 SLTTKGFQYQIVHTVIQNLLEIEKDQEELEDLIYKELDKQYQKLSKKHDQYELKQRIINAL 240
+L TKGF Y + +L ++D + EDL+ KELDKQY+KLS+K+D Y LKQ++ AL

Sbjct: 181 ALLTKGFSYDLAKHSLNHLNFDQDQNEIEDLLDKELDKQYRKLRSKYDGYTLKQKLYQAL 240

15 Query: 241 MRKGYQYEDIKSALREYL 258
RKG Y +DI LR YL

Sbjct: 241 YRKGYNDDINCKLRNYL 258

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 263

A DNA sequence (GBSx0278) was identified in *S.agalactiae* <SEQ ID 835> which encodes the amino acid
sequence <SEQ ID 836>. Analysis of this protein sequence reveals the following:

Possible site: 33
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3912(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04659 GB:AP001510 unknown conserved protein in B. subtilis
[Bacillus halodurans]

35 Identities = 96/175 (54%), Positives = 122/175 (68%)

Query: 1 MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTEENDGRWVTREP 60
M PK G I IQSYKH+GS+HR W +T+VLK T +IG ND LV E+DGR W TREP

40 Sbjct: 1 MNFPKVGSKIQIQSYKHNGSIHRIWEETIVLKGTSKVVIGGNDRILVKESDGRHWRTREP 60

Query: 61 AIVYFHKKYWFNI IAMIRETGVSYYCNLASPYILDPEALKYIDYDLDVKVFADGEKRLLD 120
AI YF + WFN I MIR G+ +YCNL +P+ D EALKYIDYDLD+KVF D +LLD

Sbjct: 61 AICYFDSEQWFNTIGMIRADGIYFYCNLGTPTWDEEALKYIDYDLDIKVFPDMTFKLLD 120

45 Query: 121 VDEYEQHKQAMNYPTDIDYILKENVKILVEWINENKGPFSYSSYINIWYKRYLELK 175
DEY H+ M YP +ID IL+ +V LV WI++ KGPFF+ ++ WY+R+L+ +

Sbjct: 121 EDEYAMHRKMMKYPPEIDRILQRSVDELVSWIHQKGPFPAPQFVESWYERFLQYR 175

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 837> which encodes the amino acid
50 sequence <SEQ ID 838>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3912(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-329-

Identities = 155/177 (87%), Positives = 165/177 (92%)

Query: 1 MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP 60
 M+LPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP
 5 Sbjct: 1 MKLPEKEDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP 60

Query: 61 AIVYFHKKYWFNIIAMIRETGVSYVCNLASPYILDPEALKYIDYDLVDVKVFADGEKRLLD 120
 AIVYFHKKYWFNIIAMIR+ GVSYYCNLASPY++D EALKYIDYDLVDVKVFADGEKRLLD
 10 Sbjct: 61 AIVYFHKKYWFNIIAMIRDNGVSYVCNLASPYMDTEALKYIDYDLVDVKVFADGEKRLLD 120

Query: 121 VDEYEQHKAQMNYPTDIDYILKENVKILVEWINENKGPFSYINIWIYKRYLELKKR 177
 VDEYE HK +M Y D+D+ILKENVKILV+WIN KGPFS +YI IWYKRYLELK R
 Sbjct: 121 VDEYETHKEMQYSADMDFILKENVKILVDWINHEKGPFSKAYITTIWYKRYLELKNR 177

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 264

A DNA sequence (GBSx0288) was identified in *S.agalactiae* <SEQ ID 839> which encodes the amino acid sequence <SEQ ID 840>. This protein is predicted to be jag protein. Analysis of this protein sequence
 20 reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.1666(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:BAB07782 GB:AP001520 spoIIJ-associated protein [Bacillus halodurans]
 Identities = 54/198 (27%), Positives = 98/198 (49%), Gaps = 6/198 (3%)

Query: 100 DVVEEYIEEVDTELEKEDVSQPELPKIDDKNVTTSEAIKIDLLPNIEVAAQVTKYVE 159
 + VE+ I E+ T E+ + E PK ++ + A+ ++ + P+ + ++E
 35 Sbjct: 13 EAVEQATIELGTTTRERITYTVVEEPKSGLFGILGSKPAVIEVVVKPD---PVDRAKAFLE 69

Query: 160 NIIYEMDLDA--TIETTTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDFRSK 217
 ++ EMD++ TIE + N+ E + G +IG G+ L SLQ L + +
 40 Sbjct: 70 ELLQEMDMEVEVTIEKDPATVLEFNISGEQ-DLGLTIGKRGQTLDLSLQYLVNLVANKEEGE 128

Query: 218 SFSVSINVHDYVEHRTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTIATIEGV 277
 + ++ +Y R E L+ +++A + L T P ++PMS ERK +H + + V
 Sbjct: 129 FIRIKLDAENYRARRKEALVQLAERLASKALRTKRPVSLEPMSAHERKIIHTALQELGDV 188

45 Query: 278 ESYSEGNDPNRFVVVTKK 295
 E+YSEG R VV+ K
 Sbjct: 189 ETYSEGQGIGRHHVVIAPK 206

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 841> which encodes the amino acid
 50 sequence <SEQ ID 842>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3721(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-330-

Identities = 176/302 (58%), Positives = 223/302 (73%), Gaps = 32/302 (10%)

Query: 23 MVLFTGATVEEAIEKGLQELNISRLRAHIKVVSRKKKGFGLGFGKKPAKVEIEGITDEVTD 82
 MVLFTG TVEEAIE GLQEL +SRL+AHIKV+S+EKKGFLGFGKKPA+V+IEGI+D+
 5 Sbjct: 1 MVLFTGKTVEEAIEETGLQELGLSRLKAHIKVISKEKKGFGLGFGKKPAQVDIEGISDKTVY 60

Query: 83 INESVALKNI-----KNVPS--SVDVVEEYIEEVDETLEKEDVSQPELPKIDDK----- 129
 + A + + +N P+ S DV E I+ + LE ED L D
 10 Sbjct: 61 KADKATRGVPEDINRQNTPAVNSADVEPEETKAT-QRLEAEDTKVPLMSSEDSFAQTPS 119

Query: 130 ---NVVITSEA-----IEKIDL-----LPNIEVAAQVTKYVENIYEMDLATI 171
 VT ++A +E+ ++ +IE AA +V+ YV IYEMD++AT+
 15 Sbjct: 120 NLAETVTETKAQQPSIPVEESEVPQDAGNDGFSKDIEKAAQEVSDYVTKIYEMDIEATV 179

Query: 172 ETTTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDRFSKFSVSVINVDYVEH 231
 ET+ ++RQINLQIETPEAGR+IGYHGKVLKSLQLLAQN+LHDR+SK+FSVS+NVHDYVEH
 20 Sbjct: 180 ETSNNRQINLQIETPEAGRVIGYHGKVLKSLQLLAQNFLHDRYSKNFSVSLNVHDYVEH 239

Query: 232 RTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTATIEGVESYSEGNDPNRFVV 291
 RTETLIDF++K+A+RVLE+ + Y MDPMSNSERK VHKT+++IEGV+SYSEGNDPNR+VV
 25 Sbjct: 240 RTETLIDFTQKVAKRVLSEGDYTMDPMSNSERKIVHKTVSSIEGVDSYSEGNDPNRYVV 299

Query: 292 VT 293
 V+
 25 Sbjct: 300 VS 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 265

30 A DNA sequence (GBSx0290) was identified in *S.agalactiae* <SEQ ID 843> which encodes the amino acid sequence <SEQ ID 844>. This protein is predicted to be 60 kd inner-membrane protein (yidC). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> May be a lipoprotein

35 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 (52 - 75)
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 (192 - 211)
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 (124 - 144)
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (167 - 184)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA78595 GB:Z14225 SpoIIJ [Bacillus subtilis]
 Identities = 79/243 (32%), Positives = 142/243 (57%), Gaps = 5/243 (2%)

Query: 1 MKKKLKTFSLILLTGSLIVACG--RGEVSSHSA TLWEQ-IVYAFAKSIQWLS--FNHSIG 55
 MK+++ ++ LL C + +++ S W++ +VY ++ I +++ + G
 50 Sbjct: 1 MKRRIGLLSMVGVMFLLAGCSSLVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYG 60

Query: 56 LGIILFTLIIRAIMPLYNMQMKSSQKMQEIQPRKELQKKYPGKDPDNRLKLNDEM QSM 115
 L IIL T++IR +++PL Q++SS+ MQ +QP +++L++KY KD + KL E ++
 55 Sbjct: 61 LSIILVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLKEKYSSKDQKTQQLQOETMAL 120

Query: 116 YKAEGVNPNYASVLFLLIQLFVLWALFQALTRVSFLKVCTFLSLELSQPDYPYILPVLAAL 175
 ++ GVN P A P+LIQ+P+L + A+ R + +FL +L + DPYIILP++A +
 60 Sbjct: 121 FQKHGVNPLAGCFPIILIGFYHAIMRTQAISEHSFLWFDLGEKDPYIILPIVAGV 180

Query: 176 FTFLSTWLINKAAVEKNIALTLMTYVMPFIILVTSFNFAAGVVLYWTVSNAFQVFQILL 235

-331-

TF+ L ++N + +M ++MP +I+V + NF + + LYW V N F + Q L+
 Sbjct: 181 ATFVQQKLLMMAGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLI 240

Query: 236 NNP 238

5

P

Sbjct: 241 KGP 243

A related GBS sequence was identified <SEQ ID 10783> which encodes amino acid sequence <SEQ ID 10784>.

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 845> which encodes the amino acid sequence <SEQ ID 846>. Analysis of this protein sequence reveals the following:

Possible site: 49

- 15 >>> May be a lipoprotein
 INTEGRAL Likelihood = -6.32 Transmembrane 198 - 214 (197 - 220)
 INTEGRAL Likelihood = -5.52 Transmembrane 59 - 75 (57 - 80)
 INTEGRAL Likelihood = -4.25 Transmembrane 130 - 146 (129 - 150)
 INTEGRAL Likelihood = -2.28 Transmembrane 173 - 189 (170 - 189)
- 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the databases:

>GP:BAA05234 GB:D26185 stage III sporulation [Bacillus subtilis]
 Identities = 90/249 (36%), Positives = 150/249 (60%), Gaps = 6/249 (2%)

- 30 Query: 16 IVPLVLLLVACG--RGEVTAQSSSGWDQ-LVYLFARAIQWLS--FDGSIGVGIIIFTLTI 70
 +V + +LL C + +TA S WD+ +VY + I +++ + G+ IIL T+ I
 Sbjct: 13 MVGVFMLLAGCSSVKEPITADSEHFWDKYVVYPLSELITYVAKLTGDNLYLSIILVTILI 72
- 35 Query: 71 RLMLPLFNMQIKSSQKMODIQPELRELQRKYAGKDTQTRMKLAEESSQALYKKYGVNPPYA 130
 RL+++PL Q++SS+ MQ +QPE+++L+ KY+ KD +T+ KL +E+ AL++K+GVNP A
 Sbjct: 73 RLLILPLMIKQLRSSKAMQALQPEMQKLEKYSKDKQKTQKQLQOETMALFQKHGVNPLA 132
- 40 Query: 131 SLLPLLIQMPVMIALFQALTRVSLKGTGFLWVELAQHDHLYLLPVLA AVFTFLSTWLTN 190
 P+LIQMP++I + A+ R + +FLW +L + D Y+LP++A V TF+ L
 Sbjct: 133 GCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYYILPIVAGVATFVQQKLM 192
- 45 Query: 191 LAAKEKNVMMITVMIYVMPLMIFFMGFNLAGSVVLYWTVSNAFQVQVQLLLNNP-FKIIE 249
 ++N M +M+++MP+MI N + + LYW V N F + Q L+ P K E
 Sbjct: 193 AGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLIKGPDIKKNPE 252
- Query: 250 RQRLANEK 258
 Q+ ++K
 Sbjct: 253 PQKAGGKKK 261

An alignment of the GAS and GBS proteins is shown below:

- 50 Identities = 172/270 (63%), Positives = 217/270 (79%), Gaps = 1/270 (0%)
- Query: 1 MKKKLKTFSLLLTGSLLVACGRGEVSSHSATLWEQIVYAFAKSIQWLSFNHSIGLGIIL 60
 +KK +K ++ L LLVACGRGEV++ S++ W+Q+VY FA++IQWLSF+ SIG+GIIL
 Sbjct: 7 VKKNIKIARIVPLV-LLLVACGRGEVTAQSSSGWDQLVYLFARAIQWLSFDGSIGVGIIIL 65
- 55 Query: 61 FTLLIRAIMMPLYNMOMKSSQKMQEIQPRKELQKKYPGKDPNRLKLNDEMOSMYKAEG 120
 FTL IR ++MPL+NMQ+KSSQKMQ+IQP L+ELQ+KY GKD R+KL +E Q++YK G
 Sbjct: 66 FTLLIRLMLPLFNMQIKSSQKMODIQPELRELQRKYAGKDTQTRMKLAEESSQALYKKYG 125
- 60 Query: 121 VNPYASVLP LLIQLPVLWALFQALTRVSLKGTGFLWVELAQHDHLYLLPVLA AVFTFLS 180
 VNPYAS+LPLLIQ+PV+ ALFQALTRVSLK GTFL +EL+Q D Y+LPVLA A+FTFLS
 Sbjct: 126 VNPYASLLPLLIQMPVMIALFQALTRVSLKGTGFLWVELAQHDHLYLLPVLA AVFTFLS 185

-332-

Query: 181 TWLTNKA AVEKNIALTLMTYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILLNNPYK 240
 TWLTN AA EKN+ +T+M YVMP +I FN ASGVVLYWTVSNAFQV Q+LLLLNNP+K
 Sbjct: 186 TWLTNLA AKEKNVMMTVMIYVMPIMIFFMGNLASGVVLYWTVSNAFQVQVQLLLLLNNPFK 245

Query: 241 IIKVREEAVRVAHEKEQRVKRAKRKASKKR 270
 II R+ E+ R +RA++KA K++
 Sbjct: 246 IIAERQRLANEKEKRLRERRARKKAMKRK 275

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8527> and protein <SEQ ID 8528> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 20 Crend: 5
 McG: Discrim Score: 4.90
 GvH: Signal Score (-7.5): -0.39
 Possible site: 42
 >>> May be a lipoprotein
 ALOM program count: 4 value: -7.38 threshold: 0.0
 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 (52 - 75)
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 (192 - 211)
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 (124 - 144)
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (167 - 184)
 PERIPHERAL Likelihood = 2.54 217
 modified ALOM score: 1.98
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

32.8/62.3% over 242aa
 Bacillus subtilis
 EGAD|17722| stage III sporulation protein j precursor Insert characterized
 OMNI|NT01BS4782 -identity Insert characterized
 SP|Q01625|SP3J_BACSU STAGE III SPORULATION PROTEIN J PRECURSOR. Edit characterized
 GP|40023|emb|CAA44401.1||X62539 unnamed protein product Insert characterized
 GP|467388|dbj|BAA05234.1||D26185 stage III sporulation Insert characterized
 GP|2636651|emb|CAB16141.1||Z99124 alternate gene name: spo0J87 Insert characterized
 PIR|I40437|I40437 stage III sporulation protein spoIIJ - Insert characterized
 ORF02221(301 - 1014 of 1413)
 EGAD|17722| S4098(3 - 245 of 261) stage III sporulation protein j precursor { acillus
 subtilis}OMNI|NT01 S4782 -identitySP|Q01625|SP3J_ ACSU STAGE III SPORULATION PROTEIN J
 PRECURSOR.GP|40023|emb|CAA44401.1||X62539 unnamed protein product { acillus
 subtilis}GP|467388|dbj| AA05234.1||D26185 stage III sporulation { acillus
 subtilis}GP|2636651|emb|CA 16141.1||Z99124 alternate gene name: spo0J87 { acillus
 subtilis}PIR|I40437|I40437 stage III sporulation protein spoIIJ - acillus subtilis
 %Match = 17.0
 %Identity = 32.8 %Similarity = 62.2
 Matches = 79 Mismatches = 88 Conservative Sub.s = 71
 219 249 279 309 339 393 420
 DFVVVIAKRGVEELDYQALEKNLIHVLKIAGLI*KGIKLKKLKTFTSLILLTGSLLVACG--RGEVSSHSA TLWEQ-IVYA
 :||::: : :: || | : :: | :||:::
 MLLKRRIGLLLSMVGVMFLLAGCSSVKEPITADSPHFWDKYVVP
 10 20 30 40
 474 504 534 564 594 624 654
 FAKSIQWLS--FNHSIGLGIILFTLITRAIMMPLYNMQMKSSQXMQEIQPRLLKELQKKYPGKDPDNRLKLNDQMOSMYKA

-333-

```

::: | ::: : || || |::|| ::|| |::||: || :|| ::||::|| || : || | ::::
LSELTIVAKLTGDNVGLSIILVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLKEYSSKDQKTQQKLQOETMALFQK
      60      70      80      90      100      110      120

5   684      714      744      774      804      834      864      894
EGVNPYASVLPPLLIQLPVLWALFQALTRVSLKVGTFSLSELSQDPDPYIILPVLAAFLSTWLTNKAAVEKNIALTLM
|||| | :|:|:|:|:| ::|:| : : :|| ::| : |||||:|:| : ||: | :| : :|
HGVNPLAGCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYIILPVIAGVATFVQQKLMAGNAQQNPQMAMM
      140      150      160      170      180      190      200

10  924      954      984      1014      1044      1074      1104      1134
TYVMPFIILVTSFNFASGVVLYWTVSNAFQVFOILLNNPYKLIKVRREAVRVAHEKEQVRVKRAKRKASKKRRK*ENHGII
::|| :|:| : || : : ||| | | : | :| : : : :
LWIMPIMIIVFAINPPAALSPLYWVVGNLFMIAQTFLIKGPDIKKNPEPQKAGGKKK
      220      230      240      250      260

15

```

Example 266

A DNA sequence (GBSx0291) was identified in *S.agalactiae* <SEQ ID 847> which encodes the amino acid sequence <SEQ ID 848>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3778 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9579> which encodes amino acid sequence <SEQ ID 9580> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

30 >GP:CAA44400 GB:X62539 homologous to E.coli rnpA [Bacillus subtilis]
    Identities = 52/109 (47%), Positives = 77/109 (69%), Gaps = 1/109 (0%)

Query: 21 LKKTYRVKSDKDFQMFISRGKNVANRKFVIYYLEK-EQKHFRVGVISVSKKLGNAVVRNAI 79
      LKK R+K ++DFQ +F G +VANR+FV+Y L++ E RVG+SVSKK+GNAV+RN I
35 Sbjct: 4 LKKRNRLKKNEDFQKVFKHGTSTVANRQFVLYTLTDPENDELRVGLSVSKKIGNAVVRNRI 63

Query: 80 KRKIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGL 128
      KR IR L +K L++ D+++IARK +L Y+ +K+L H+ + + L
40 Sbjct: 64 KRLIRQAFLEEKERLKEKDYIIIRKAPASQLTYEETKKSLOHLFRKSSL 112

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 849> which encodes the amino acid sequence <SEQ ID 850>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3820 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
50      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

      Identities = 73/109 (66%), Positives = 88/109 (79%)

Query: 21 LKKTYRVKSDKDFQMFISRGKNVANRKFVIYYLEKEQKHFRVGVISVSKKLGNAVVRNAIK 80
      LKKTYRVK +KDFQ IF GK+ ANRKFVIY+L + Q HFRVGVISV KK+GNAV RNA+K
55 Sbjct: 1 LKKTYRVKREKDFQAIKDGKSTANRKFVIYHLNRQDHFVRVGVISVGKIGNAVTRNAVK 60

Query: 81 RKIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGLI 129
      RKIRHV+++ L+ DFVVIARKGV L+YQ L++NL HVLK+A L+

```

Sbjct: 61 RKIRHVIMALGHQLKSEDFVVIARKGVHSLEYQELQQNLHHVLKLAQLL 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 267

A DNA sequence (GBSx0292) was identified in *S.agalactiae* <SEQ ID 851> which encodes the amino acid sequence <SEQ ID 852>. This protein is predicted to be glycerol-3-phosphate dehydrogenase, NAD-dependent (gpsA). Analysis of this protein sequence reveals the following:

10 Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1429(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8529> which encodes amino acid sequence <SEQ ID 8530> was also identified. There is a signal peptide at residues 1-19. The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAA86746 GB:U32164 NAD(P)H-dependent dihydroxyacetone-phosphate
reductase [Bacillus subtilis]
Identities = 177/333 (53%), Positives = 241/333 (72%)

25 Query: 18 QKIAVLGPGSWGITALAQVLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSKIK 77
+K+ +LG GSWGITALA VL DNG+EV +W + + I +IN H N+ Y ++ L + IK
Sbjct: 2 KKVIMLGAGSWGITALALVLTNDNGNEVCVWAHRADLIHQINELHENKDYLPNVKLSTSIK 61

Query: 78 YTNLEEAINNVDSILFVVPVKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLSTILE 137
T+++EA+++ D I+ VPTK R V +Q + K V +H SKG+EP + R+S I+E
30 Sbjct: 62 TTDMEKAVSDADVIVAVPTKAIAREVLROAVPFITKAVFVHVSKEIEPDSLRLRISEIME 121

Query: 138 EEISEQYRSDIVVVGPSHAEEAIVRDITLITAASKDIEAAKYVQKLFSNHYFRLYTNTD 197
E+ R DIVV+SGPSHAEE +R T +TA+SK + AA+ VQ LF NH FR+YTN D
35 Sbjct: 122 IELPSDVRRDIVVLSGPSHAEEVGLRHATTVTASSKSMRAAEEVQDLFINHNFRVYTNPD 181

Query: 198 VVGVEITAGALKNIIVAGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFSGLS 257
++GVE GALKNIIA+ AG GLGYGDNAKAA+ITRGLAEI RLG ++G +PLTFSGL+
Sbjct: 182 IIGVEIGGALKNIIAALAGITDGLGYGDNAKAAIITRGLAEIARLGTMGGNPLTFSGLT 241

40 Query: 258 GVGDLI VTGTSVHSRNRWAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNINVM 317
GVGDLI VT TSVHSRNRWAG+ LG+G KLED+ + MGMV+EG+ TTK AY++++ +V M
Sbjct: 242 GVGDLI VTCTSVHSRNRWAGNLLGKGYKLEDVLEEMGMVVEGVRTTKAAYQLSKKYDVKM 301

45 Query: 318 PITEAIYKSIYEGANIKDSILDMSNEFRSENE 350
PITEA+++ ++ G ++ ++ +M+ E E
Sbjct: 302 PITEALHQVLENGQKVETAVEESLMARGKTHEME 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 853> which encodes the amino acid sequence <SEQ ID 854>. Analysis of this protein sequence reveals the following:

50 Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0882(Affirmative) < succ>
55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-335-

Identities = 287/338 (84%), Positives = 316/338 (92%)

Query: 15 MTKQKIAVLGPGSWGTTALAQVLNDNGHEVRLWGNVVEQIEINTNHTNQRYFKDITLDSK 74
 MTKQK+A+LGPWSGTAL+QVLNDNGH+VRLWGN+ +QIEEINT HTN+ YFKDI LD
 5 Sbjct: 1 MTKQKVAILGPGSWGTTALSQVLNDNGHDVRLWGNIPDQIEEINTKHTNRHYFKDIVLDKN 60

Query: 75 IKAYTNLEEAINNVDISILFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPFGTHERLST 134
 I A +L +A+++VD++LFVVPTKVTRLVA+QVA +L HKVV+MHASKGLEP THERLST
 10 Sbjct: 61 ITATLDLGOALSDVDAVLVFPVPTKVTRLVARQVAAILDHKVVVMHASKGLEPETHERLST 120

Query: 135 ILEEEISEQYRSDIVVSGPASHAEAIVRDITLITAASKDIEAAKYVQKLFNSHYFRLYT 194
 ILEEEI +RS++VVVSGPASHAE IVRDITLITAASKDIEAAKYVQ LFSNHYFRLYT
 15 Sbjct: 121 ILEEEIPAFRSEVVVSGPASHAETIVRDITLITAASKDIEAAKYVQSLFSNHYFRLYT 180

Query: 195 NTDVVGVTAGALKNI IAVGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFS 254
 NTDV+GVETAGALKNI IAVGAGALHGLGYGDNAKAA+ITRGLAEITRLGV+LGADPLT+S
 20 Sbjct: 181 NTDVIGVTAGALKNI IAVGAGALHGLGYGDNAKAAVITRGLAEITRLGVKLADPLTYS 240

Query: 255 GLSGVGDLIVTGTSVHSRNWRAGDALGRGEKLEDIEKNMGVIEGISTTKVAYETIAQNLN 314
 GLSGVGDLIVTGTSVHSRNWRAG ALGRGEKLEDIE+NMGMVIEGI+TTKVAYETIAQ+L
 25 Sbjct: 241 GLSGVGDLIVTGTSVHSRNWRAGAALGRGEKLEDIERNMGMVIEGIATTKVAYETIAQDLG 300

Query: 315 VYMPITEAIYKSIYEGANIKDSILDMSNEFRSENEWH 352
 VYMPIT AIYKSIYEGA+IK+SIL MMSNEFRSENEWH
 30 Sbjct: 301 VYMPITTAIYKSIYEGADIKESILGMSNEFRSENEWH 338

SEQ ID 8530 (GBS291) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 5; MW 38.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 2; MW 64kDa).

30 GBS291-GST was purified as shown in Figure 226, lane 10-11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 268

A DNA sequence (GBSx0293) was identified in *S.agalactiae* <SEQ ID 855> which encodes the amino acid sequence <SEQ ID 856>. This protein is predicted to be glucose-1-phosphate uridylyltransferase (gtaB). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have a cleavable N-term signal seq.

40 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA28714 GB:AB001562 glucose-1-phosphate uridylyltransferase
 [Streptococcus mutans]
 Identities = 263/296 (88%), Positives = 285/296 (95%)

50 Query: 2 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR 61
 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR
 Sbjct: 5 KVRKAVIPAAGLGTRFLPATKALAKEMLPIDVKPTIQFIVEEALKSGIEDILVVTGKSKR 64

55 Query: 62 SIEDHFDSNFELEYNLKEKGKNEILLVDETT GIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
 SIEDHFDSNFELEYNL++KGK +LLKL++TT I LHFIRQSHPRGLGDAVLQAKAFVGN
 Sbjct: 65 SIEDHFDSNFELEYNLQKGKTDLLKLVDITTAINLHFIRQSHPRGLGDAVLQAKAFVGN 124

Query: 122 EFPVVMGLDGLMDITNNKVIPLTKQLINDFEATHASTIAVMEVPHEDVSAYGVIAPOGEG 181

-336-

EPPFVVMGLGDDLMDIT++K IPLT+QL+ND+E THASTIAVMEVPHEVDVSAYGVIA PQGEG
 Sbjct: 125 EPPFVVMGLGDDLMDITDDKAIPLTRQLMNDYEE THASTIAVMEVPHEVDVSAYGVIA PQGEG 184
 Query: 182 VNGLYSVNTFVEKPSPEEAPS NLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
 5 V+GLYSV+TFVEKP+P+EAPS NLAIIGRYLLTPEIF ILETQ+PGAGNE+QLTDAIDTLN
 Sbjct: 185 VSGLYSVDTFVEKPAPKEAPS NLAIIGRYLLTPEIFTILETQEPGAGNEVQLTDAIDTLN 244
 Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKTS 297
 KTQRVFAR+F G RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII+LGK L++ S
 10 Sbjct: 245 KTQRVFAREFKGRYDVGDKFGFMKTSIDYALKHPQVKEDLKAYIIELGKKLDQKS 300

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 857> which encodes the amino acid sequence <SEQ ID 858>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 257/295 (87%), Positives = 277/295 (93%)

25 Query: 2 KVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKR 61
 KVRKA+IPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIE+ILVVTGK+KR
 Sbjct: 3 KVRKAIIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEEILVVTGKAKR 62
 Query: 62 SIEDHFDSNFELEYNLKEKGKNELKLVDETT GIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
 30 SIEDHFDSNFELEYNL+ KGKNELKLVDETT I LHFIRQSHPRGLGDAVLQAKAFVGN
 Sbjct: 63 SIEDHFDSNFELEYNLQAKGKNELKLVDETTAINLHFIRQSHPRGLGDAVLQAKAFVGN 122
 Query: 122 EPPFVVMGLGDDLMDITNNKVIPLTKQLINDFEATHASTIAVMEVPHEVDVSAYGVIA PQGEG 181
 EPPFVVMGLGDDLMDITN PLTKQL+ D++ THASTIAVM+VPHEVDVS+YGVIA PQG+
 35 Sbjct: 123 EPPFVVMGLGDDLMDITNASAKPLTKQLMEDYDKTHASTIAVMKVPHEVDVSSYGVIA PQGKA 182
 Query: 182 VNGLYSVNTFVEKPSPEEAPS NLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
 V GLYSV+TFVEKP PE+APS+LAIIGRYLLTPEIF ILE Q PGAGNE+QLTDAIDTLN
 Sbjct: 183 VKGLYSVDTFVEKPPQPEDAPSDLAIIGRYLLTPEIFGILERQTPGAGNEVQLTDAIDTLN 242
 40 Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKT 296
 KTQRVFAR+F G+RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII LGK+LEK+
 Sbjct: 243 KTQRVFAREFKGNRYDVGDKFGFMKTSIDYALEHPQVKEDLKNYIIEKLGKALEKS 297

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 269

A DNA sequence (GBSx0294) was identified in *S.agalactiae* <SEQ ID 859> which encodes the amino acid sequence <SEQ ID 860>. Analysis of this protein sequence reveals the following:

50 Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 (27 - 45)
 ----- Final Results -----
 55 bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-337-

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
[Bacillus subtilis]
Identities = 148/346 (42%), Positives = 222/346 (63%), Gaps = 16/346 (4%)

5 Query: 31 LTLTSLSVLTTLTACGNRSDKSAN---KSDIKVAMVTNQGVDKSFNQSAWEGLOKWK 87
++L+ + L ACGN S + K+ VAMVT+ GGVDKSFNQSAWEG+Q +GK+
Sbjct: 1 MSLVIAAGTILGACGNSEKSSGSGEGKKNFSVAMVTDVGGVDKSFNQSAWEGIQAFGKE 60

10 Query: 88 KGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKDVKYV 146
GL KG NG+DY QS +++D+ NL+ A ++LI+G+G+ + D+I ++++ K+ +
Sbjct: 61 NGLKKGKNGYDYLQSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

15 Query: 147 IVDDIIKGENVASVTFADNEAAYLAGVAAAKTTKTVTGFIGGMEGVVVKRFEAGFKAG 206
I+D ++ K+NVAS+TF + E ++L GVAAA ++K+ +GF+GGME ++K+FE GF+AG
Sbjct: 121 IIDAVVD-KDNVASITPKQEGRSFLVGVAALSSKSGKIGFVGGMESELIKFEVGFRAG 179

20 Query: 207 VKSIDPAIKVAVSYAGSFTDAAGKGTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENET 266
V++++P V V YAG F A GK A + Y +GVDVIY +AG TG G+F+EAK
Sbjct: 180 VQAVMPKAVVEVKYAGGFDKADVGKATAESMYKSGVDVIYHAGATGTGVFTAK---NL 236

25 Query: 267 RKESNK--VWVIGVDRDQSQEGNYVSKDGKKANFVLASTIKEVGKSLQSVAELETKKQYP 324
+KE K VWVIGVD+DQ EG +G N L S +K+V ++ V + ++P
Sbjct: 237 KKEDPKRDVWVIGVDRDQYABGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFP 293

Query: 325 GGVTVFVGLKDSGVDI--KEHQLSSEGSVAVKKAKEDIVSGKIQVP 368
GG+ +GL GV I + LS + AV K K+ I+ G +++P
Sbjct: 294 GGETLTYGLDQDGVGISPSKQNLSDVDIKAVDKWKKKIIDG-LEIP 338

There is also homology to SEQ ID 862.

30 A related GBS gene <SEQ ID 8531> and protein <SEQ ID 8532> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 20 Crend: 3
Sequence Pattern: CGNR
SRCFLG: 0

35 McG: Length of UR: 19
Peak Value of UR: 2.31
Net Charge of CR: 2
McG: Discrim Score: 5.09
GvH: Signal Score (-7.5): -3.29

40 Possible site: 19
>>> May be a lipoprotein
Amino Acid Composition: calculated from 21
ALOM program count: 0 value: 5.20 threshold: 0.0
PERIPHERAL Likelihood = 5.20 90

45 modified ALOM score: -1.54

*** Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 52.8/73.9% over 239aa

Listeria

monocytogenes

SP|Q48754| CD4+ T CELL-STIMULATING ANTIGEN PRECURSOR. Insert characterized

60 GP|7240601|gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen Insert characterized

ORF02225(385 - 1086 of 1710)

SP|Q48754|TCSA_LISMO(8 - 247 of 268) CD4+ T CELL-STIMULATING ANTIGEN

PRECURSOR.GP|7240601|gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen {Listeria

monocytogenes}

-338-

```

%Match = 21.7
%Identity = 52.7 %Similarity = 73.8
Matches = 125 Mismatches = 59 Conservative Sub.s = 50

5   294      324      354      384      414      444      465      489
    NFLWEK*NKVC*MIFLCYDRNFLFLCDYNLLGGSFVSNRKIIGLTLLSLSVLTLTACGNRSD--KSANKS--DIKVAMVT
                                :|::: | : | ||: || | :|| | ||||
                                MKKRTFALALSMIIASGVILGACGSSSDDKKSSDDKSSKDFTVAMVT
                                10      20      30      40

10  519      549      579      606      636      666      696      726
    NQGGVDDKSFNQSAWEGQLQKWGKKGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKD
    : ||||:|||||||:| : || :||:|:|:|:|: ||:| | :||:|:|: | | ||:|:
    DTGGVDDRSFNQSAWEGQLQKFGKANDMEKGTGYNLQSASEADYKTNLNTAVRSDYDLIYGIGYKLKDAIEEVSKQKPK

15      60      70      80      90      100      110      120

    756      786      816      846      876      906      936      966
    VKYIVDDI IKGKENVASVTFADNEAAYLAGVAAAKTTTKTKTVGFIGGMEGVVVKRFEAGFKAGVKSIDPAIKVAVSYAG
    : |||| | :|| | : ||:|:| || | |||| ||||:|:| | : |||| | |||:|:| | |||
    NQFAIVDDTIDDRDNVVSIGFDNDGSYLVGVVAGLITTKTNKVGFGVGGVKGTVIDRFEAGFTAGVKAVNPNAQIDVQYAN

20      140      150      160      170      180      190      200

    996      1026      1056      1086      1116      1146      1176      1206
    SFTDAAGKKTIAATQYATGVVDVIYQAAGGTGAGIFSEAKTENETRKESNKNVWVIGVDRDQSQEGNYVSKDGKKANFVLAS
    | | ||: ||: |::|||:|:|||| |::||| : : :
    DFAKADKQQTASSMYSSGVVDVIFHAAGGTGNGVFAEAKNLKKKDLQMPYGNKSLGCGFGG
    220      230      240      250      260

```

A related GBS nucleic acid sequence <SEQ ID 10947> which encodes amino acid sequence <SEQ ID 10948> was also identified.

SEQ ID 8532 (GBS108) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 7; MW 39.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 9; MW 64.6kDa).

The GBS108-GST fusion product was purified (Figure 202, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 273), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 270

A DNA sequence (GBSx0295) was identified in *S.agalactiae* <SEQ ID 863> which encodes the amino acid sequence <SEQ ID 864>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -12.74    Transmembrane  206 - 222 ( 197 - 224)
    INTEGRAL    Likelihood = -3.72    Transmembrane  174 - 190 ( 171 - 194)
    INTEGRAL    Likelihood = -3.19    Transmembrane   98 - 114 (  98 - 116)
    INTEGRAL    Likelihood = -1.54    Transmembrane  120 - 136 ( 120 - 139)
    INTEGRAL    Likelihood = -0.90    Transmembrane  157 - 173 ( 157 - 173)

50  ----- Final Results -----
        bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-339-

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
Identities = 126/218 (57%), Positives = 166/218 (75%)

5 Query: 8 KEYPTTVLLVSLTTLVFLMLQLTYGSQAESSQVIFQFGGIQGDYLKAYPTNLWRLISPIF 67
KE P T +S+T L+F++MQ+ YGS A+S QV+FQFGG+ G +K+ P+ LWRL++PIF
Sbjct: 5 KEKPVTFFFLSVTILLFIVMQVFYGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

10 Query: 68 VHIGWEHFLNLGLALYFVGQMGESIWGSLRFLILYLISGLMGNIFTLFFTPHVVAAGAST 127
+HIGWEHFL+N L LYFVGQ+ ESTWGS FL+LY+LSG+MGN+ TLFFTPHVVAAGAST
Sbjct: 65 IHIGWEHFLINSLTLYFVGQLAESIWGSRFFLLLYVLSGIMGNVLTTLFFTPHVVAAGAST 124

15 Query: 128 SLFGVFSATAIAGYFGKNPYLKQVGKSYQVMILLNFFNIPTPGVSLAGHVGGVGLV 187
SLFG+F+AI + GYFG N LK +GKSYQ +I+LNL N+F P V + GH+GG +GG L
Sbjct: 125 SLFGLFAAIVVVGYPFGHNQLKSIGKSYQTLIILNLVMNLFMPNVGIVGHLGGALGGALA 184

Query: 188 AIFLTKQNGSLLFKTWQSILALMIFIIVSISLIGLSLV 225
A+FL + LF Q AL+ ++ +++ LI LSL+
Sbjct: 185 AVFLPTLLDAELFTKKQKTSALLSYLTLALVLITLSLM 222

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 865> which encodes the amino acid sequence <SEQ ID 866>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -9.92 Transmembrane 214 - 230 (212 - 232)
INTEGRAL Likelihood = -5.36 Transmembrane 135 - 151 (128 - 153)
INTEGRAL Likelihood = -1.81 Transmembrane 101 - 117 (100 - 117)
INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (182 - 199)
INTEGRAL Likelihood = -0.53 Transmembrane 166 - 182 (166 - 182)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
Identities = 72/128 (56%), Positives = 94/128 (73%)

40 Query: 106 FLLLYVLSGVMGNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGKSYQT 165
FLLLYVLSG+MGN T + TP VAAGASTSLFGLFAAIVV+ + G NQ LK +GKSYQT
Sbjct: 95 FLLLYVLSGIMGNVLTTLFFTPHVVAAGASTSLFGLFAAIVVVGYPFGHNQLKSIGKSYQT 154

45 Query: 166 LIVVNLLMNLFMPNVSMAGHIGGVGGALLSIVFPTKMRVITVKTKRMLALVSYGIILV 225
LI++NL+MNLFMPNV + GH+GG +GGAL ++ PT + K ++ AL+SY + +
Sbjct: 155 LIILNLVMNLFMPNVGIVGHLGGALGALAAVFLPTLLDAELFTKKQKTSALLSYLTLAL 214

Query: 226 GVLVLGFL 233
++ L +
Sbjct: 215 VLITLSLM 222

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/132 (47%), Positives = 92/132 (68%)

55 Query: 94 GSLRFLILYLISGLMGNIFTLFFTPHVVAAGASTSLFGVFSATAIAGYFGKNPYLKQVGK 153
G FL+LY+LSG+MGN FT + TP VAAGASTSLFG+F+AI + + GKN LK +GK
Sbjct: 102 GLTPFLLLYVLSGVMGNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGK 161

60 Query: 154 SYQVMILLNLFNIPTPGVSLAGHVGGVGLVLAIFLTKQNGSLLFKTWQSILALMIFI 213
SYQ +I++NL N+F P VS+AGH+GG+VGG L++I + K + +LAL+ +
Sbjct: 162 SYQTLIVVNLLMNLFMPNVSMAGHIGGVGGALLSIVFPTKMRVITVKTKRMLALVSYG 221

Query: 214 IVSISLIGLSLV 225
I+ + ++ L +
Sbjct: 222 IILVGVLVLGFL 233

-340-

A further corresponding DNA sequence was identified in *S.pyogenes* <SEQ ID 9083> which encodes the amino acid sequence <SEQ ID 9084>. Analysis of this protein sequence reveals the following:

```

5       Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -7.70    Transmembrane    12 - 28 ( 7 - 30)

10     ----- Final Results -----
          bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS sequences follows:

```

15     Score = 74.5 bits (180), Expect = 5e-16
      Identities = 37/96 (38%), Positives = 48/96 (49%)

      Query: 1  MTQLLKRYPPXXXXXXXXXXXXXAMQVVYGHLATGAQAIYQVGGMFGLLVKAMPDQLWRL 60
                M + K YP                      MQ+ YG A +Q I+Q GG+ G +KA P LWRL
20     Sbjct: 3  MKKFAKEYPTTVLLVSLTTLVFLLMQLTYGSAESSQVIFQFGGIQGDYLKAYPTNLWRL 62

      Query: 61  VTPXXXXXXXXXXXXXVNGLTLYFVQGIVEDLWGSRLF 96
                ++P                      +NGL LYFVGQ+ E +WGS F
25     Sbjct: 63  ISPIFVHIGWEHFLNGLALYFVGQMGESIWGSLRF 98

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 271

A DNA sequence (GBSx0296) was identified in *S.agalactiae* <SEQ ID 867> which encodes the amino acid sequence <SEQ ID 868>. Analysis of this protein sequence reveals the following:

```

30     Possible site: 22
>>> Seems to have no N-terminal signal sequence

35     ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40     >GP:BAA28715 GB:AB001562 hypothetical protein [Streptococcus mutans]
      Identities = 96/173 (55%), Positives = 129/173 (74%)

      Query: 1  MEKKLLRKEVLITLKSQPQAYKSEVDCKLLEAFIKTKAYQNSCVIATYLSFDYEYNTQLL 60
                M KK R +V+ LK Q +A K D +LLE I+ +AYQ + VIATYL+F +E++T LL
45     Sbjct: 1  MMKKDYRTQVIEDLKKQDKAKKVLRDEQLLEELIQLEAYQKAHVIATYLAFFPEFDTSL 60

      Query: 61  IKQALCDGKRVLVPKTYPKGKMFVDYQKDNLRTPFGLLPEVNDRAVEKASIDLIHVP 120
                I+QA D K ++VPKTYP+ KMIFV Y + +L+ T FGL EP ++ A+EK++IDLIHVP
50     Sbjct: 61  IEQAQRDNKSIVVPKTYPQRKMIFVVYDEADLQITKFLKEPRSEEALEKSAIDLIHVP 120

      Query: 121 LIFNNKGFRIGYGAGYFDRYLSDFEGDTISTIYRCQRQDFVEEKHDVAVKEVL 173
                L FNN+G+RIG+GAGY+D+YL+DF+GDT+STIY Q+ F D+ VKEVL
      Sbjct: 121 LAFNNEGYRIGFGAGYYDQYLDADFQGDITVSTIYSFQQFTFEPSFFDIPVKEVL 173

```

55 A related GBS nucleic acid sequence <SEQ ID 10925> which encodes amino acid sequence <SEQ ID 10926> was also identified.

-341-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 272

- 5 A DNA sequence (GBSx0297) was identified in *S.agalactiae* <SEQ ID 869> which encodes the amino acid sequence <SEQ ID 870>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.44    Transmembrane  161 - 177 ( 161 - 177)
10    INTEGRAL    Likelihood = -0.22    Transmembrane   29 - 45 ( 28 - 45)

----- Final Results -----
          bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9305> which encodes amino acid sequence <SEQ ID 9306> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
    [Streptococcus mutans]
    Identities = 344/401 (85%), Positives = 374/401 (92%)

Query: 1 MDLPENYDKKEEFSRIQKAAEKIKSDSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAE 60
25 ++LP+NYDKKEEF+RI+KAAEKIKSDSEVLVVIGIGGSYLGAA+AAIDFLN+ F NL+ EE
Sbjct: 49 LNLFPQNYDKKEEFARIKAAEKIKSDSEVLVVIGIGGSYLGARAAIDFLNSSFVNLENKEE 108

Query: 61 RKAPQILYAGNSISSYTLADLVEIVYQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 120
30 RKAPQILYAGNSISS YLADLV+YV DK+FSVNVISKSGTTTEPAIAFRVFK+LLVKKYG
Sbjct: 109 RKAPQILYAGNSISSNYLADLVYVADKDFS NVISKSGTTTEPAIAFRVFKDLLVKKYG 168

Query: 121 QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNGGRFSVLTAVGLLPAAASGADIT 180
QEEAN+RIYATTD+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPAAASGAD+
35 Sbjct: 169 QEEANQRIYATTD RVKGAVKVEADANGWETFVVPDSVGGRTVLTAVGLLPAAASGADLD 228

Query: 181 ALMEGANAAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 240
LM GA AAR+D SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQL
Sbjct: 229 QLMAGAEAAARQDYSSAELENEAYQYAAIRNLYRKGYVTEVLANYEPSLQYFSEWWKQL 288

Query: 241 AGESEKGDQKGIYPTSANFSTDLSLGGQFIQEGYRNLFFETVVRVEKPRKNVTIPELTEDL 300
AGESEKGDQKGIYPTSANFSTDLSLGGQFIQEG RNLFFETV+RVEK RKN+ +PE EDL
40 Sbjct: 289 AGESEKGDQKGIYPTSANFSTDLSLGGQFIQEGNRNLFETVIRVEKARKNIIIVPEAAEDL 348

Query: 301 DGLGYLQGDVDFVNKKATDGVLLAHTDGGVPMFVTLPTQDAYTLGYTIYFFELAIGLS 360
DGL YLQGDVDFVNKKATDGVLLAHTDGGVPM F+T+P QD +TLGY IYFFELAIGLS
45 Sbjct: 349 DGLAYLQGDVDFVNKKATDGVLLAHTDGGVPMNTFLTIPQDEFTLGVTYIYFFELAIGLS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELS AELNARL 401
GYLN VNPFDQPGVEAYK+NMFALLGKPGFEEL AELNARL
50 Sbjct: 409 GYLNGVNPFDQPGVEAYKKNMFALLGKPGFEELGAELNARL 449

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 871> which encodes the amino acid sequence <SEQ ID 872>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
55 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.44    Transmembrane  209 - 225 ( 209 - 225)

```

-342-

INTEGRAL Likelihood = -0.22 Transmembrane 77 - 93 (76 - 93)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
[Streptococcus mutans]
Identities = 369/449 (82%), Positives = 408/449 (90%)

Query: 1 MSHITFDYSKVLSEFAGQHEIDFLQGQVTEADKLLREGTGFGSDFLGWLDLPENYDKDEF 60
M+HI FDYSKVL F HE+D++Q QVT AD+ LR+GTGFG++ GWL+LP+NYDK+EF
15 Sbjct: 1 MTHIKFDYSKVLGKFLASHELDYIQMQVTADEALRKTGFGAEMTGWLNLPQNYDKKEEF 60

Query: 61 ARILTAAEKIKADSEVLVVGIGGSYLGAKEAIDFLNHHFANLQTAKERKAPQILYAGNS 120
ARI AAEKIK+DSEVLVVGIGGSYLGA+AAIDFLN F NL+ +ERKAPQILYAGNS
20 Sbjct: 61 ARIKKAEEKIKSDSEVLVVGIGGSYLGAKEAIDFLNSSFVNLENKEERKAPQILYAGNS 120

Query: 121 ISSTYLADLVEYVQDKFSNVISKSGTTTEPAIAFRVFKELLVKKYQBEANKRIYATT 180
ISS YLADLV+YV DK+FSNVISKSGTTTEPAIAFRVFK+LLVKKYQBEAN+RIYATT
Sbjct: 121 ISSNYLADLVYVADKDFSNVISKSGTTTEPAIAFRVFKDLLVKKYQBEANQRIYATT 180

25 Query: 181 DKVKGAVKVEADANNWETFFVVPDNGGRFVLTAVGLLPAAAGADITALMEGANAPARKD 240
D+VKGAVKVEADAN WETFFVVPD+VGGRF+VLTAVGLLPAAAGAD+ LM GA AAR+D
Sbjct: 181 DRVKGAVKVEADANGWETFFVVPDSVGGRFVLTAVGLLPAAAGADLDQLMAGAEAAARQD 240

Query: 241 LSSDKISENIAYQYAAVRNVLYRKGYYTEILANYEPSLQYFGEWWKQLAGESEGKDQKGI 300
SS ++SEN AYQYAA+RN+LYRKG+Y+TE+LANYEPSLQYF EWWKQLAGESEGKDQKGI
30 Sbjct: 241 YSSAELSENEAYQYAAIRNLYRKGYYTEVLANYEPSLQYFSEWWKQLAGESEGKDQKGI 300

Query: 301 YPTSANFSTDHLHSLGQFIQEGYRNLFFETVIRVDNPRKNVILPELAEDLDGLYLQGDVD 360
YPTSANFSTDHLHSLGQFIQEG RNLFFETVIRV+ RKN+++PE AEDLDGL YLQGDVD
35 Sbjct: 301 YPTSANFSTDHLHSLGQFIQEGNRNLFFETVIRVEKARKNVLPEAAEDLDGLAYLQGDVD 360

Query: 361 FVNKKATDGVLLAHTDGGVPMNFVTLPAQDEFTLGYYIYFFELAIYVSGYMNVAVNPFDQP 420
FVNKKATDGVLLAHTDGGVPM F+T+P QDEFTLG YIYFFELAI +SGY+N VNPFDQP
40 Sbjct: 361 FVNKKATDGVLLAHTDGGVPMNFTLTPQDEFTLGYYIYFFELAIYVSGYLNVAVNPFDQP 420

Query: 421 GVEAYKRNMFALLGKPGFEALS AELNARL 449
GVEAYK+NMFALLGKPGFE L AELNARL
Sbjct: 421 GVEAYKRNMFALLGKPGFEELGAELNARL 449

45 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus
uberis]
Identities = 58/91 (63%), Positives = 69/91 (75%)

50 Query: 6 KRPITIFLLGLTGLIFIAMQVYVGHATGAQAIYQVGGMFGLLVKAMPDQLWRLVTPIF 65
K P+T F L +T L+FI MQV YG A Q ++Q GGMFGL+VK+MP QLWRLVTPIF
Sbjct: 5 KEKPVTFFFLSVTILLFIVMQVPYGSWAKSPQVVFQFGGMFGLLVKSMPSQLWRLVTPIF 64

Query: 66 IHIGFGHFFVNGLTLYFVGQIVEDLWGSRLF 96
IHIG+ HF +N LTLYFVGQ+ E +WGSR F
55 Sbjct: 65 IHIGWEHFLINSLTLYFVGQLAESIWGSRLF 95

An alignment of the GAS and GBS proteins is shown below:

Identities = 380/401 (94%), Positives = 392/401 (96%)

60 Query: 1 MDLPENYDKKEEFSRIQKAAEEKIKSDSEVLVVGIGGSYLGAKEAIDFLNHHFANLQTAEE 60
+DLPENYDK+EF+RI AAEKIK+DSEVLVVGIGGSYLGAKEAIDFLN+HFANLQTA+E
Sbjct: 49 LDLPENYDKDEFARILTAAEEKIKADSEVLVVGIGGSYLGAKEAIDFLNHHFANLQTAKE 108

-343-

Query: 61 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 120
 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG
 Sbjct: 109 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 168

5 Query: 121 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPD NVGGRFSVLTAVGLLPIAASGADIT 180
 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPD NVGGRFSVLTAVGLLPIAASGADIT
 Sbjct: 169 QEEANKRIYATTDKVKGA VKVEADANNWETFVVPD NVGGRFSVLTAVGLLPIAASGADIT 228

10 Query: 181 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 240
 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL
 Sbjct: 229 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 288

15 Query: 241 AGESEKDKQKGIYPTSANFSTDLHSLGQFIQEGYRNL FETVVRVEKPRKNVTIPELTEDL 300
 AGESEKDKQKGIYPTSANFSTDLHSLGQFIQEGYRNL FETV+RV+ PRKNV IPEL EDL
 Sbjct: 289 AGESEKDKQKGIYPTSANFSTDLHSLGQFIQEGYRNL FETVIRVDNPRKNVTIPELAEDL 348

20 Query: 301 DGLGYLQGKD VDFVNKKATDGVLLAHTDGGVPMFVTLPTQDAYTLGYTIYFFELAIGLS 360
 DGLGYLQGKD VDFVNKKATDGVLLAHTDGGVPMFVTL P QD +TLGYTIYFFELAI +S
 Sbjct: 349 DGLGYLQGKD VDFVNKKATDGVLLAHTDGGVPMFVTLPAQDEFTLGYTIYFFELAIAVS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELS AELNARL 401
 GY+N+VNPFDQPGVEAYKRNMFALLGKPGFE LS AELNARL
 Sbjct: 409 GYMNAVNPFDQPGVEAYKRNMFALLGKPGFEALS AELNARL 449

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 273

A DNA sequence (GBSx0298) was identified in *S.agalactiae* <SEQ ID 873> which encodes the amino acid sequence <SEQ ID 874>. Analysis of this protein sequence reveals the following:

30 Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 654 - 670 (653 - 671)
 INTEGRAL Likelihood = -1.65 Transmembrane 113 - 129 (113 - 129)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9463> which encodes amino acid sequence <SEQ ID 9464> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]
 45 Identities = 536/864 (62%), Positives = 663/864 (76%), Gaps = 3/864 (0%)

Query: 20 ETTDVALAIDTLVQNGLKALDEMR--QLNQEQVDYIVAKASVAALDAHGELALHAVEETG 77
 +T V I+ LV+ AL E + QE++DYIV KASVAALD H LA AVEETG
 Sbjct: 5 QTMTVDEHINQLVRKAQVALKEYLKPEYEQEKIDYIVKASVAALDQHCALAAA AVEETG 64

50 Query: 78 RGVFEDKATKNLFACEHVVMNMRHTKTIVGVIEEDDVTGLTLIAEPVGVVCGITPTTNPTS 137
 RG+FEDKATKN+FACEHV + MRH KTVG+I D + G+T IAEPVGVVCG+TP TNPTS
 Sbjct: 65 RGIFEDKATKNI FACEHVTHEMRHAKTVGI INVDEL YGITEIAEPVGVVCGVTFVTNPTS 124

55 Query: 138 TAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCQWIEQPSIDAT 197
 TAIFKSLIS+KTRNPI+F+FHPSA + S AA+IVRDAAIAAGAPENC+QWIE I+A+
 Sbjct: 125 TAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIVRDAAIAAGAPENCQWIEFGGIEAS 184

60 Query: 198 NALMNHGDIATILATGGNAMVKAAYS CGKPALGVGAGNVPAVVEKSANIRQA AHDIVMSK 257
 N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+VMSK

-344-

Sbjct: 185 NKL MNHPGVATILATGGNAMVKAAYSSGKPALGVGAGNVPTIYEKTCNIKQAANDVVMMSK 244

Query: 258 SFDNGMVCASEQAVIIDKEYIKEFVEEFKSYHTYFVNKKEKALLEEFCGAKANSKNCAG 317
SFDNGM+CASEQA IIDKEYI + VEE K+ YF+N++EKA LE+F FG A S +

5 Sbjct: 245 SFDNGMICASEQAIIIDKEYIDQVVEEMKTLGAYFTINEEEKAKLEKFMFGVNAYSADVNN 304

Query: 318 AKLNPNIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAVLKAEST 377
A+LNP G S W AEQ G VPE NI+ A C EV EPLTREKLSPV+A+LKAE+T

10 Sbjct: 305 ARLNPKCPGMS PQWFAEQVGIVKPEDCNIICAVCKEVPNEPLTREKLSPVLAAILKAENT 364

Query: 378 EDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIGDVYNAF 437
+DG++KA MVEFNG GHSAAIH+ D + ++ ++A R++ N+PS+ GGIG +YN

Sbjct: 365 QDGIDKAEAMVEFNNGRHSAAIHSNDKAVVEKYALTMKACRILHNTPSQGGIGSIYNYI 424

15 Query: 438 LPSLTGCGSYGRNSVGDNVSAINLLNIKKVRRRNNMQNFKVPSKTYFERDSIQYLQKC 497
PS TLGCGSYG NSV NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL +

Sbjct: 425 WPSFTLGCGSYGNSVSANVTYHLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYLAEL 484

Query: 498 RDVERVMIVTDHAMVELGFLDRIEQDLRRNKVVVYQIFAEVEPDPPDITTMKGTDLMRT 557
+++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPD I TV KG +M T

20 Sbjct: 485 KELS KIFIVSDRMMYKLG YVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAVMNT 544

Query: 558 FKPDITIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKEPPELGKTKFVA 617
F PD IIA+GGGS MDAK+MWL YE PE DF + QKF+D+RKRAFKEP +GKK + +

25 Sbjct: 545 FGPDNIIAIGGGSAMDAKIMWLLYEHPEADFFAMKQKFIDLRKRAFKEPPTMGKKARLIC 604

Query: 618 IPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAADTGMDV 677
IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP M++P ADTG+DV

30 Sbjct: 605 IPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLDV 664

Query: 678 LTHATEAYVSQMANDYTDGLALQAIKIVFDYLERSVKDADFEAREKMHNASTMAGMAFAN 737
L HATEAYVS MAN+YTDGLA +A+K+VF+ L +S + D EAREKMHNA+T+AGMAFA+

Sbjct: 665 LVHATEAYVSVMANEYTDGLARBAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMAFAS 723

35 Query: 738 AFLGISHSMAHKIGAQFHTVHCRNTNAILLPYVIRYNGTRPAKTATWPKYNYRADEKYQD 797
AFLG+ HSMHAK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++Y +

Sbjct: 724 AFLGMDHSMHAKVGAAPHLPHGRCVAVLLPHVIRYNGQKPRKLAMWPKYNFYKADQRYME 783

Query: 798 IAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNFRDQGIDEKEWKEKSRELAFLAYED 857
+A+++GL TP E VE++AKA +L F+ IDE W K E+A LA+ED

40 Sbjct: 784 LAQMVLKCNTPAEGVEAFKACEELMKATETITGFKKANIDEAAMSKVPEMALLAFED 843

Query: 858 QCSPANPRLEPMVDHMQEIIIDAYY 881
QCSPANPR+PMV M++I++ AYY

45 Sbjct: 844 QCSPANPRVPMVKDMEKILKAAYY 867

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 875> which encodes the amino acid sequence <SEQ ID 876>. Analysis of this protein sequence reveals the following:

Possible site: 55

50 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -3.66	Transmembrane	643 - 659 (642 - 660)
INTEGRAL	Likelihood = -1.81	Transmembrane	102 - 118 (102 - 118)

55 ----- Final Results -----

bacterial membrane	---	Certainty=0.2466(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

60 The protein has homology with the following sequences in the databases:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]

Identities = 535/870 (61%), Positives = 669/870 (76%), Gaps = 3/870 (0%)

65 Query: 6 NIVETTSVSVITDALVQKGLAALAEEMRKLD--QEVDYIVAKASVAALDAHGELAKHAYE 63

-345-

+T +T +V I+ LV+K AL+E K + QE++DYIV KASVAALD H LA A E
 Sbjct: 2 STQQTMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAAAVE 61

5 Query: 64 ETGRGVFEDKATKHLFACEHVNNMRHQKTVGIIEDDDVTGLTLIAEPVGVICGITPTTN 123
 ETGRG+FEDKATK++FACEHV + MRH KTVGII D + G+T IAEFVGVC+CG+TP TN
 Sbjct: 62 ETGRGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPVGVVCGVTPVTN 121

10 Query: 124 PTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAATAAGAPENCQVWVETPSL 183
 PTSTAIFKSLIS+KTRNPI+F+FHPSA + S AA+IVRDAATAAGAPENC+QW+E +
 Sbjct: 122 PTSTAIFKSLISIKTRNPIVFSFHPALKCSIMAAKIVRDAATAAGAPENCQWIEFGGI 181

15 Query: 184 EATNALMNHGDIATILATGGNAMVKAAYS CGKPALGVGAGNVPAYVEKSANIRQAAHDIV 243
 EA+N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+V
 Sbjct: 182 EASNKL MNHPGVATILATGGNAMVKAAYS SGPALGVGAGNVPYIEKTCNIKQAANDVV 241

20 Query: 244 MSKSFNDNGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKANSKN 303
 MSKSFNDNGM+CASEQA IIDKEIYD V E K+ YF+N++EKA LE+F PG A S +
 Sbjct: 242 MSKSFNDNGMICASEQAIIIDKEIYDQVVEEMKTLGAYFINEBEKAKLEKFMFGVNAYSAD 301

25 Query: 304 CAGAKLNPNIIVGKIPATWIAEQAGFTVPEGTNIIAABCKEVSENEPLTREKLSPIVAVLKS 363
 A+LNP G W AEQ G VPE NI+ A CKEV NEPLTREKLSPV+A+LK+
 Sbjct: 302 VNNARLNP KCPGMS PQWFAEQVGIVKVPEDCNIICAVCKE VGPNEPLTREKLSPIVAILKA 361

30 Query: 364 ESREDGVEKARQMVEFNGLGHSAAIHTADAE LAKEFGTRIRAIRVIWNSPSTFGGIGDVY 423
 E+ +DG++KA MVEFNG GHSAAIH+ D + +++ ++A R++ N+PS+ GGIG +Y
 Sbjct: 362 ENTQDGDIDKAEAMVEFNGRGHSAAIHSNDKAVVEKYALTMKACRILHNTPSQGGIGISTY 421

35 Query: 424 NAFLEPSLTLCGSGYGRNAVGDVNSAINLLNIKKVGRRRNNMOWFKVPSKTYFERDSIQYL 483
 N PS TLGCGSYG N+V NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL
 Sbjct: 422 NYIWPSTLTCGSGYGGNSVSANVTYHNNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYL 481

40 Query: 484 QKCRDVERVMIVIDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPTITVMKGTEL 543
 + +++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPD I TV KG +
 Sbjct: 482 AELKELSKIFIVSDRMMYKLGIVYDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAV 541

45 Query: 544 MRTFKPDITIIALGGGSPMDAAKVMWLFYEOPEVDFHDLVQKFMDIRKRAFKEPFGKKTG 603
 M TF PD IIA+GGGS MDAAK+MWL YE PE DF + QKF+D+RKRAFKEP +GKK +
 Sbjct: 542 MNTFGPDNIIAIGGGSAMDAKIMWLLYEHPEADFFAMKQKFDLRKRAFKEPFTMGKKAR 601

50 Query: 604 FVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPGFIAADTG 663
 + IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP +++P ADTG
 Sbjct: 602 LICIPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSAIVDPMFTMSLPKRAIADTG 661

55 Query: 664 MDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTIADFEAREKMHNASTMAGMA 723
 +DVL HATEAYVS MAN++TDGLA +A+K+VF+NL KS D EAREKMHNA+T+AGMA
 Sbjct: 662 LDVLVHATEAYVSVMANEYTDGLAREAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMA 720

60 Query: 724 FANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRDEK 783
 FA+AFLG+ HMAHK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++
 Sbjct: 721 FASAFLGMDHSMHAKVGAAPHLPCHRCVAVLLPHVIRYNGQKPRKLANWPKYNFYKADQR 780

65 Query: 784 YQDIAKLLGLPASTPBEAVESYAKAVYDLGCRVGIQMNFAQGIDENEWEKHSRELAYLA 843
 Y ++A+++GL +TP E VE++AKA +L FK IDE W E+A LA
 Sbjct: 781 YMELAQMVGLKCNTPAEGVEAFKACEELMKATETITGFKKANIDEAAWMSKVPEMALLA 840

70 Query: 844 YEDQCSPANRPLPMVDHMQEIIEDAYGYA 873
 +EDQCSPANPR+PMV M++I++ AYY A
 Sbjct: 841 FEDQCSPANPRVPMVKDMEKILKAAYYPIA 870

60 An alignment of the GAS and GBS proteins is shown below:

Identities = 827/880 (93%), Positives = 852/880 (95%)

65 Query: 12 MTEKTKAVETTTDALAIDTLVQNGLKALDEMRLNQEQVDYIVAKASVAALDAHGELALH 71
 MTE VETT V++ ID LVQ GL AL+EMR+L+QEQVDYIVAKASVAALDAHGELA H
 Sbjct: 1 MTEGHNTVETTSVSVTIDALVQGLAALDEMRKLDQEQVDYIVAKASVAALDAHGELAKH 60

Query: 72 AVEETGRGVFEDKATKNLFACEHVNNMRHTKTVGVIEEDDDVTGLTLIAEPVGVVCGITP 131

-346-

A EETGRGVFEDKATK+LFACEHVNNMRH KTVG+IEEDDVTGLTLIAEPVGV+CGITP
 Sbjct: 61 AYEETGRGVFEDKATKHLFACEHVNNMRHQKTVGIIEEDDVTGLTLIAEPVGVICGITP 120
 Query: 132 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCQWIEQ 191
 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCQW+E
 Sbjct: 121 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCQWVET 180
 Query: 192 PSIDATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 251
 PS++ATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH
 Sbjct: 181 PSLEATNALMNHGDIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 240
 Query: 252 DIVMSKSFNDGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKAN 311
 DIVMSKSFNDGMVCASEQAVIIDKEIY +FV EFKSYHTYFVNKKEKALLEEFCFGAKAN
 Sbjct: 241 DIVMSKSFNDGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKAN 300
 Query: 312 SKNCAGAKLNPNIIVGKSAVWIAEQAGFTVPEGTNIIAAECTEVSEKEPLTREKLSFVIAV 371
 SKNCAGAKLNPNIIVGK A WIAEQAGFTVPEGTNIIAAEC EVSE EPLTREKLSFVIAV
 Sbjct: 301 SKNCAGAKLNPNIIVGKPATWIAEQAGFTVPEGTNIIAAECKEVSENEPLTREKLSFVIAV 360
 Query: 372 LKAESTEDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIG 431
 LK+ES EDGVEKARQMVEFNGLGHSAAIHT DA+LA+EFGTRIRAIRVIWNSPSTFGGIG
 Sbjct: 361 LKSESREDGVEKARQMVEFNGLGHSAAIHTADAELAKEFGTRIRAIRVIWNSPSTFGGIG 420
 Query: 432 DVYNAFLPSLTGCGSYGRNSVGDNVSAINLLNIIKKVGRRRNNMQWFKVPSKTYFERDSI 491
 DVYNAFLPSLTGCGSYGRN+VGDNVSAINLLNIIKKVGRRRNNMQWFKVPSKTYFERDSI
 Sbjct: 421 DVYNAFLPSLTGCGSYGRNAVGDNVSAINLLNIIKKVGRRRNNMQWFKVPSKTYFERDSI 480
 Query: 492 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPIITVMKG 551
 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPIITVMKG
 Sbjct: 481 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPIITVMKG 540
 Query: 552 TDLMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKEPFGK 611
 T+LMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKEPFGK
 Sbjct: 541 TELMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKEPFGK 600
 Query: 612 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAA 671
 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALV+TVPGFIAA
 Sbjct: 601 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPFGFIAA 660
 Query: 672 DTGMDVLTHATEAYVSQMANDYTDGLALQAIKIVFDYLSVSKDADFEAREKMHNASTMA 731
 DTGMDVLTHATEAYVSQMAND+TDGLALQAIKIVFD LE+SVK ADFEAREKMHNASTMA
 Sbjct: 661 DTGMDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMA 720
 Query: 732 GMAFANAFGLGISHMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA 791
 GMAFANAFGLGISHMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA
 Sbjct: 721 GMAFANAFGLGISHMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYRA 780
 Query: 792 DEKYQDIAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNFRDQGIDEKEWKEKSRELA 851
 DEKYQDIAKLLGLPA+TPEEAVESYAKAVYDLG R+GI+MNF+ QGIDE EWKE SRELA
 Sbjct: 781 DEKYQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMNFKAQGIDENEWKEKSRELA 840
 Query: 852 FLAYEDQCSPANRPLPMVDHMQEIIEDAYYGYEERPGRRK 891
 +LAYEDQCSPANRPLPMVDHMQEIIEDAYYGY ERPGRRK
 Sbjct: 841 YLAYEDQCSPANRPLPMVDHMQEIIEDAYYGYAERPGRRK 880

A related GBS gene <SEQ ID 8533> and protein <SEQ ID 8534> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -4.68
 GvH: Signal Score (-7.5): -2.48
 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -2.66 threshold: 0.0
 INTEGRAL Likelihood = -2.66 Transmembrane 100 - 116 (99 - 117)
 PERIPHERAL Likelihood = 3.61 173
 modified ALOM score: 1.03

-347-

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 SEQ ID 8534 (GBS432) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 7; MW 41kDa).

GBS432-GST was purified as shown in Figure 223, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 274

A DNA sequence (GBSx0299) was identified in *S.agalactiae* <SEQ ID 877> which encodes the amino acid sequence <SEQ ID 878>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

20

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3444(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 880.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 275

A DNA sequence (GBSx0300) was identified in *S.agalactiae* <SEQ ID 881> which encodes the amino acid sequence <SEQ ID 882>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

35 INTEGRAL Likelihood = -8.39 Transmembrane 74 - 90 (69 - 94)
 INTEGRAL Likelihood = -5.31 Transmembrane 168 - 184 (163 - 186)
 INTEGRAL Likelihood = -4.83 Transmembrane 34 - 50 (29 - 52)
 INTEGRAL Likelihood = -0.75 Transmembrane 202 - 218 (202 - 219)

40

----- Final Results -----

 bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA17305 GB:AL021926 hypothetical protein Rv0111 [Mycobacterium tuberculosis]

Identities = 70/218 (32%), Positives = 104/218 (47%), Gaps = 12/218 (5%)

50

Query: 9 VRITGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKTKKIDFVSFCRRR 68
 +R + LVL H GGF+GVD FF SGFLIT+LL+DE +T +ID F RR

-348-

Sbjct: 39 LRATAVALVLASHGGIPGMGGGFIGVDAPFVLSGFLITSLLLDELGRITGRIDLSGFWIRR 98

Query: 69 FYRIFPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGNYESQFI-P 127
 R+ P LVLMLV L + S + A +T+N+ + +Y +Q P

5 Sbjct: 99 ARRLLPALVLMVLTVSAARALFPDQALTGLRSDAIAAFLWTANWRFVAQNTDYFTQGAPP 158

Query: 128 HLFVHTWSLSIEVHFYVLWGL----TVWLLSKRSKDQKQLRGTLFLISMGI FGVSFLTMTF 183
 HTWSL +E +YV+W L LL+ R++ ++ R T+ + F ++ L

10 Sbjct: 159 SPLQHTWSLGVVEEQYYVWVPLLLIGATLLLAARAR-RRCRATVGGVRFAAFLIASLGT 217

Query: 184 VRAFFVDNFST-----IYFSTLSHIFPFFLGAMVATI 215
 A F++ IYF T + +G+ A +

Sbjct: 218 ASATAAVAFSAATRDRIYFGDTTRAQALLIGSAAAAL 255

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 879> which encodes the amino acid sequence <SEQ ID 880>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have an uncleavable N-term signal seq

20	INTEGRAL	Likelihood = -10.83	Transmembrane	325 - 341 (313 - 346)
	INTEGRAL	Likelihood = -9.29	Transmembrane	237 - 253 (234 - 258)
	INTEGRAL	Likelihood = -7.91	Transmembrane	166 - 182 (162 - 188)
	INTEGRAL	Likelihood = -6.10	Transmembrane	72 - 88 (68 - 92)
	INTEGRAL	Likelihood = -4.09	Transmembrane	264 - 280 (260 - 281)
	INTEGRAL	Likelihood = -2.87	Transmembrane	371 - 387 (370 - 390)
25	INTEGRAL	Likelihood = -2.66	Transmembrane	34 - 50 (32 - 50)
	INTEGRAL	Likelihood = -1.91	Transmembrane	3 - 19 (3 - 19)
	INTEGRAL	Likelihood = -0.85	Transmembrane	136 - 152 (136 - 154)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 167/226 (73%), Positives = 195/226 (85%)

Query: 1 MRIKWFSLVRIITGLLLVLLYHFFKNSFPGGFVGVDFIIFTFSGFLITALLIDEFSKTKKID 60
 MRIKWFS VR+TGLLLVLLYHFFKN FPGGF+GVDFIIFTFSG+LITALLIDE++K + ID

40 Sbjct: 1 MRIKWFSFVRVTGLLLVLLYHFFKNVFPGGFIGVDIIFTFSGYLITALLIDEYTKKESID 60

Query: 61 FVSFCRRRFYRIPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGN 120
 + F +RRFYRI PPLVLM+L+TIPF FL+K DF A+IGSQI LGFT+N YEILTG +

45 Sbjct: 61 IIGFLKRRFYRIVPPLVLMILLTIPFTFLIKKDFIANIGSQITAVLGFTTNIYEILTGS 120

Query: 121 YESQFIPHLFVHTWSLSIEVHFYVLWGLTVWLLSKRSKDQKQLRGTLFLISMGI FGVSFL 180
 YESQFIPHLFVHTWSL+IEVHFY+ WG+ VWLL++R + QKQLRG LFLIS+GIF +SFL

50 Sbjct: 121 YESQFIPHLFVHTWSLAIEVHFYLFWGVFVWLLARRKETQKQLRGLLFLISLGIFAISFL 180

Query: 181 TMFVRAFFVDNFSTIYFSTLSHIFPFFLGAMVATISGIREITGRFK 226
 +MF+R+F NFS IYFS+LSH PFFLGAM ATI+GI E T RF+

Sbjct: 181 SMFIRSFMTSNFSLIYFSSLSHSFPFFLGAMFATITGINETTVRFQ 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 276

A DNA sequence (GBSx0302) was identified in *S.agalactiae* <SEQ ID 883> which encodes the amino acid sequence <SEQ ID 884>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have a cleavable N-term signal seq.

60

-349-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

!GB:AE004818 hypothetical protein [Pseudomonas aerug...
 !GB:AE004818 hypothetical protein [Pseudomonas aerug...
 !GB:AE004818 hypothetical protein [Pseudomonas aerug...
 10 !GB:AE004818 hypothetical protein [Pseudomonas aerug...
 !GB:AE004818 hypothetical protein [Pseudomonas aerug...
 >GP:AAG07403 GB:AE004818 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 33/80 (41%), Positives = 50/80 (62%)
 15
 Query: 45 KYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVHGWSTYTGDFKKQPDGQ 104
 +Y G +V+ + G+G+L Y+NG +Y G F +G+ G GT+ G Y+G F G DGQ
 Sbjct: 39 RYRGELVDGRLEGQGRLLDYDNGAWYAGRFEHGLLHGHTWQAGDSRYSGGFAAGLFDGQ 98
 20
 Query: 105 GRLNAKNNKVKYKGTFFKQGIY 124
 GRL + VY+G F+QG++
 Sbjct: 99 GRLAMADGSVYQGGFRQGLF 118
 Identities = 31/91 (34%), Positives = 46/91 (50%), Gaps = 2/91 (2%)
 25
 Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVHGWSTY 93
 QG YD G Y G + + G G +G Y G F G+F+G+G G Y
 Sbjct: 52 QGRLDYDNGAW-YAGRFEHGLLHGHTWQAGDSRYSGGFAAGLFDGQGRLAMADGSVYQ 110
 30
 Query: 94 GDFKKQPDGQGRNALNNKVKYKGTFFKQGIY 124
 G F++G DG+G L + + Y+G F++G+Y
 Sbjct: 111 GGFRQGLFDGEGSLEQQGTR-YRGGFRKGLY 140
 Identities = 31/91 (34%), Positives = 42/91 (46%), Gaps = 1/91 (1%)
 35
 Query: 32 SSQGVFSYDGGKIKYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVHGWST 91
 S QG G +Y GS + G+G + G+ Y G F +G GKG + G
 Sbjct: 141 SQGTLDGSDGS-RYQGSFRQGRLEGEFSFSDSQGNQYAGTFRDGLNGKGRWSPDGD 199
 Query: 92 YTGDFKKQPDGQGRNALNNKVKYKGTFFKQ 122
 Y G FK Q GQGR + + V+ G F +G
 40 Sbjct: 200 YVGQPKDNQPHGQGRYESASGDVWIGRFSEG 230
 Identities = 31/91 (34%), Positives = 45/91 (49%), Gaps = 4/91 (4%)
 Query: 34 QGVFSYDGGK----IKYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVH 89
 QG+F +G +Y G +G+G L +G Y+G F G EG+G+F G
 45 Sbjct: 115 QGLFDGEGSLEQQGTRYRGGFRKGLYSGQGTLDGSDGSRYQGSFRQGRLEGEFSFSDSQ 174
 Query: 90 WSYTGDFKKQPDGQGRNALNNKVKYKGTFFK 120
 Y G F+ GQ +G+GR + + Y G FK
 50 Sbjct: 175 NQYAGTFRDGLNGKGRWSPDGDRIYVQFK 205
 Identities = 28/87 (32%), Positives = 45/87 (51%), Gaps = 1/87 (1%)
 Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVHGWSTY 93
 +G FS G +Y G+ + + GKG+ + +GD Y G F + F G+G + S G +
 55 Sbjct: 166 EGSFSDSQGN-QYAGTFRDGLNGKGRWSPDGDRIYVQFKDNQPHGQGRYESASGDVWI 224
 Query: 94 GDFKKQPDGQGRNALNNKVKYKGTFFK 120
 G F +G +G G L + Y+G F+
 Sbjct: 225 GRFSEGALNGFCELLGADGSRYRGGFQ 251
 60 Identities = 28/89 (31%), Positives = 43/89 (47%), Gaps = 2/89 (2%)
 Query: 34 QGVFSYDGGKIKYVGSIVNHHMTGKGKLTIEYNGDYKGFVNGVFEGKGTFFSVHGWSTY 93
 QG + G +Y G G+G L + G Y+G F G++ G+GT G Y
 Sbjct: 98 QGRLAMADGSV-YQGGFRQGLFDGEGSLE-QQGTRYRGGFRKGLYSGQGTLDGSDGSRYQ 155
 65
 Query: 94 GDFKKQPDGQGRNALNNKVKYKGTFFKQ 122
 G F++G+ +G+G + Y GTF+ G
 Sbjct: 156 GSFRQGRLEGEFSFSDSQGNQYAGTFRD 184

-350-

Identities = 25/80 (31%), Positives = 37/80 (46%)

Query: 45 KYVGSIVNHHMTGKGLTYENGDIYKGFVNGVFEGKGTFSVHGWSYTGDFKKGQPDGQ 104

+YVG ++ G+G+ +GD + G F G G G + G Y G F+ + GQ

Sbjct: 199 RYVGQFKDNQFHGQGRYESASGDVWIGRFSEGLNGPGEELGADGSRYRGGFQFWRFHGQ 258

Query: 105 GRLNAKNNKVVYKGTFFKQGIY 124

G L + Y+G F G Y

Sbjct: 259 GLLBQLDGTRYEGGFAAGAY 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 885> which encodes the amino acid sequence <SEQ ID 886>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -13.16 Transmembrane 20 - 36 (12 - 41)

----- Final Results -----

bacterial membrane --- Certainty=0.6265(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA16606 GB:D90899 hypothetical protein [Synechocystis sp.]

Identities = 37/89 (41%), Positives = 49/89 (54%), Gaps = 6/89 (6%)

Query: 48 KGRMHYT-----GYVINHKMNGEGKLVYPNGDIYEGTFKDGLEFEGKGTFTAKTGWLYNG 101

KG YT G V+ ++NG GK Y NCD YEGT K+G +G+G F G Y G

Sbjct: 141 KGTPIYTNGDRCSGTVVQGEINSGKCEYNNQDQYEGTLKNGQPDGEGIFRFAAGGEYEG 200

Query: 102 EFHKQANGKGVKAKNNKVVYKGTFFKQGI 130

EF G+ +G+G N ++G FKQG+

Sbjct: 201 EFQSGEFSGQGTTRIFANGNRFOGQFQKQGL 229

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/126 (53%), Positives = 93/126 (72%)

Query: 1 MKNFKITRTHLEILSLIIIVFGLSVFTLTSSQGVFSYDGGKIKYVGSIVNHHMTGK GK 60

+K + ITR LEI+S+I+I+V +SVF++ S++ +YD G++ Y G ++NH M G+GK

Sbjct: 8 VKKWSITRAKLEIVSVIVILVCAISVFSVRISNKTSLTVDKGRMHYTGYVINHKMNGEGK 67

Query: 61 LTYENGDIYKGFVNGVFEGKGTFSVHGWSYTGDFKKGQPDGQGRNLNAKNNKVVYKGTFFK 120

L Y NCD Y+G F +G+FEKGTTF + GW Y G+F KQK +G+G L AKN KVKG FK

Sbjct: 68 LVYPNGDIYEGTFKDGLEFEGKGTFTAKTGWLYNGEFHKQANGKGVKAKNNKVVYKGTFFK 127

Query: 121 QGIYQK 126

QGI+QK

Sbjct: 128 QGIFQK 133

SEQ ID 884 (GBS139) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 3; MW 13kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 2; MW 38.2kDa), in Figure 24 (lane 7; MW 38kDa) and in Figure 33 (lane 7; MW 38.2kDa).

The GBS139-GST fusion product was purified (Figure 200, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 287), which confirmed that the protein is immunoaccessible on GBS bacteria.

-351-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 277

A DNA sequence (GBSx0303) was identified in *S. agalactiae* <SEQ ID 887> which encodes the amino acid sequence <SEQ ID 888>. This protein is predicted to be holliday junction dna helicase ruvB (ruvB). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4386(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB75331 GB:Y15896 RuvB protein [Bacillus subtilis]
Identities = 196/322 (60%), Positives = 254/322 (78%)

20 Query: 3 RFLDSDAMGDEELVERTLRPOYLREYIGQDKVKDQLKIFIEAAKLDESLEDHVLLFGPPG 62
R + S+A E ++E++LRPQ L +YIGQ KVK+ L++FI+AAK+R E+LDHVLL+GPPG
Sbjct: 4 RLVSSEADNHESVIEQSLRPQNLAQYIGQHKVKENLRVFIIDAAKMRQETLDHVLLYGPPG 63

25 Query: 63 LGKTTMAFVIANELGVNLKQTSGPATIEKSGDLVAILNDLEPGDVLFIIDEIHRMPMAVEEV 122
LGKTT+A ++ANE+GV L+ TSGPAIE+ GDL AIL LEPGDVLFIIDEIHR+ ++EEV
Sbjct: 64 LGKTTLASIVANEMGVNELRTTSGPATIERPGDLAAILTALEPGDVLFIIDEIHRRLRSIEEV 123

30 Query: 123 LYSAMEDFYIDIMIGAGETSRSVHLDLPPFTLIGATTAGMLSNPLRAREFGITGHMEYYE 182
LY AMEDF +DI+IG G ++RSV LDLPPFTL+GATTR G+L+ PLR RFG+ +EYY
Sbjct: 124 LYPAMEDFCLDIVIGKGPSARSVRLDLPPFTLVGATTTRVGLLTAPLRDRFGVMSRLEYTT 183

35 Query: 183 ENDLTEIERTADIFEMKITTYEAASELARRSRGTPRIANRLKRVRDYQIMGDGLIDN 242
+ +L +I+ RTAD+FE++I +A E+ARRSRGTPR+ANRLL+RVRD+AQ++GD I ++
Sbjct: 184 QEELADIVTRTADVFEVEIDKPSALEIARRSRGTPRVANRLRRVRDFAQVLGDSRITED 243

40 Query: 243 ITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTLVSNIAEERDTVEDMYEPEYL 302
I+ AL L VD GLD++D K+L MIE +NGGPVGL T+S I EE T+ED+YEPEYL
Sbjct: 244 ISQNALERLQVDRLLGLDHIDHKLMLMGMIKFNNGGPVGLDTISATIGESHTIEDVYEPEYL 303

45 Query: 303 IQKGFIMRTRTRGRVATVKAYEH 324
+Q GFI RT GR+ T Y H
Sbjct: 304 LQIGFIQRTPRGRIVTPAVYHH 325

A related GBS nucleic acid sequence <SEQ ID 10943> which encodes amino acid sequence <SEQ ID 10944> was also identified.

50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 889> which encodes the amino acid sequence <SEQ ID 890>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0686(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below:

Identities = 282/327 (86%), Positives = 306/327 (93%)

-352-

Query: 1 MTRFLDSDAMGDEELVERTLRPQYLREYIGQDKVKDQLKIFIEAAKLDESLDHVLLFGP 60
M R LD++ MG+EE +RTLRPQYL EYIGQDKVK+Q IFIEAAK RDESLDHVLLFGP
Sbjct: 25 MARILDNNVMGNEEFSDRTLRPQYLHEYIGQDKVKEQFAIFIEAAKRRDESLDHVLLFGP 84

5 Query: 61 PGLGKTTMAFVIANELGVNLKQTSGLPAIEKSGDLVAILNDLEPGDVLFIIDEIHRMPMAVE 120
PGLGKTTMAFVIANELGVNLKQTSGLPA+EK+GDLVAILN+LEPGD+LFIDEIHRMPM+VE
Sbjct: 85 PGLGKTTMAFVIANELGVNLKQTSGLPAVEKAGDLVAILNELEPGDILFIIDEIHRMPMSVE 144

10 Query: 121 EVLYSAMEDFYIDIMIGAGETSRSVHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 180
EVLYSAMEDFYIDIMIGAG+TSRS+HLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY
Sbjct: 145 EVLYSAMEDFYIDIMIGAGDTSRSHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 204

15 Query: 181 YEENDLTEIERTADIFEMKITVEAASELARRSRGTPRIANRLKRVRDYQIMGDGLID 240
Y+E DLTEI+ERTA IFE+KI +EAA +LA RSRGTPRIANRLKRVRDYQI+GDG+I
Sbjct: 205 YQEKDLTEIVERTATIFEIKIDHEAARKLACRSRGTPRIANRLKRVRDYQIIGDGIIT 264

20 Query: 241 DNITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTLNVIAEERDTEVEDMYEP 300
ITD+ALTMLDVD EGLDY+DQKILRTMIEMY GGPVGLGTLNVIAEER+TVE+MYEP
Sbjct: 265 AQITDRALTMLDVDREGLDYIDQKILRTMIEMYQGGPVGLGTLNVIAEERNTEVEEMYP 324

Query: 301 YLIQKGFIMRTRTGRVATVKAYEHLGY 327
YLIQKGF+MRTRTGRVAT KAY HLG
Sbjct: 325 YLIQKGFIMRTRTGRVATQKAYRHLGY 351

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 278

A DNA sequence (GBSx0304) was identified in *S.agalactiae* <SEQ ID 891> which encodes the amino acid sequence <SEQ ID 892>. Analysis of this protein sequence reveals the following:

30 Possible site: 43
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.87 Transmembrane 157 - 173 (157 - 174)
INTEGRAL Likelihood = -1.49 Transmembrane 205 - 221 (205 - 222)

35 ----- Final Results -----
bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 893> which encodes the amino acid sequence <SEQ ID 894>. Analysis of this protein sequence reveals the following:

45 Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/303 (42%), Positives = 202/303 (65%)

55 Query: 1 MLKHFGSKVRNLRVTRNITREDFCGDETELSVRQLARIESGQSIPNLTAKHYIAKQLNVK 60
ML+HFG KV+ LR+ + I+RED CGDE+ELSVRQLARIE GQSIP+L+K +IAK LNV
Sbjct: 1 MLEHFGGKVKVLRLEKRISREDLCGDESELSVRQLARIELGQSIPSLSKVIFIAKALNV 60

Query: 61 LDILTGGESLELPKRYKELKYLILRIPTYADAERLKLRECQFDHIFEEFYDNLPEDECLA 120
+ LT G LELPKRYKELKYLILR PTY D +L++RE QFD IFE++YD LPE+E +

-353-

Sbjct: 61 VGYLTGADLELPKRYKELKYLIIRTPTYMDDGKLQVREEQFDEIFEDYDKLPBEEKII 120

Query: 121 IDSLQAKFEVYQTGDINFGVEVLCECFDKVKYKEKYTLNDLIIDLFILTCVVSKFNNRA 180
ID LQA + + + NFG+++L E F+++K K ++ NDLI+++L+L + + +

5 Sbjct: 121 IDCLQATLDTLLSENTNFGIDLLQEYFNQIKTKVFRQNDLILLELYLAYLDIEGMDGQY 180

Query: 181 FTKEVFQTIICKTLISQNHKLTAEEDLFWFNHVLNCFVGLCLNSEECLAEMLEVSRTMV 240
K + ++ L Q + ++LF N +++ + L N + L + +E+S++ M

10 Sbjct: 181 SDKIFYDSLLDNLSEQFEQFELDELFIWNKIIDISSLSLKNRLDNLEKAIEMSQKIMA 240

Query: 241 STHDFHKMPLYFMYQWKYFITIDNDIKSAENAYQQSIMFSKMIDDKHLIKKLELEWQEDI 300
D+++MP+ + +WKYF+ DI AE ++ + +F++M D++L KL EW++D+

Sbjct: 241 KIQDWNRMPIKLIEWKYFLIKQKDIKAEQSFMAQLFAQMTADQYLENKLIQEWKDV 300

15 Query: 301 TGH 303
+
Sbjct: 301 KSY 303

20 SEQ ID 892 (GBS319) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 4; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 7; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 279

25 A DNA sequence (GBSx0305) was identified in *S.agalactiae* <SEQ ID 895> which encodes the amino acid sequence <SEQ ID 896>. This protein is predicted to be adenylosuccinate lyase (purB). Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3358(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04344 GB:AP001509 adenylosuccinate lyase [Bacillus halodurans]
Identities = 326/430 (75%), Positives = 366/430 (84%)

40 Query: 1 MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
MIERY+RP EM AIWTEEN+Y+AWLEVEI+A EAWAELGEIPKEDV KIRE A FD++RIL
Sbjct: 1 MIERYTRPEMGAIWTEENRYQAWLEVEIVACEAWAELGEIPKEDVKKIREHASFDVERIL 60

45 Query: 61 EIEQDTRHDVVAFTTRAVSETLGEERKWHYGLTSTDVVDTA YLYKQANDIIRRDLENF 120
EIEQ+TRHDVVAFTTRAVSETLGEERKWHYGLTSTDVVDTA YL KQAN+II DL F
Sbjct: 61 EIEQETRHDVVAFTTRAVSETLGEERKWHYGLTSTDVVDTALS YLLKQANEII EADLVRF 120

Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPTTFGLKLATWYSEMKNRIERFEHAAAGVEAGKI 180
+I+ +KA EHK+T+MMGRTHGVHAEPTTFGLKLA WY EMKRN+ERF AA GV GK+

50 Sbjct: 121 LDILKEKALEHKYTVMMGRTHGVHAEPTTFGLKLALWYEMKRNLERFRLAEGVRVGKL 180

Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEIISTQVLPRLDHA EYFAVLASIATSIERMATEI 240
SGAVG +ANI PFVEQYVC+KLG+ ISTQ L RD HA EY A LA IATGIE+ A EI

55 Sbjct: 181 SGAVGTYANIDPFVEQYVCEKGLERAPISTQTLQRDRHAEYMATLALIASIEKFAVEI 240

Query: 241 RGLQKSEQREVEEFFAKGQKSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300
RGLQKSE REVEE+FAKGQKSSAMPHKRNPIGSENMTG+ARV+RGHM+ AYENV LWHE

Sbjct: 241 RGLQKSETREVEEYFAKGQKSSAMPHKRNPIGSENMTGIARVVIRGHMLAAYENVPLWHE 300

-354-

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFNGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
 RDISHSSAERII PD TI I+YMLNRFNGNIVKNLTVFPENM RNM T+GLIYSQRV+L
 Sbjct: 301 RDISHSSAERIILPDATIAINYMLNRFNGNIVKNLTVFPENMKRNMRTYGLIYSQRVLLS 360

5 Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQBEIDELFNPIYYTKR 420
 LI+KGM REEAYDLVQPK +W+ V F+ L+E++ ++TS L+ EEI+ F+ ++ K
 Sbjct: 361 LIDKGMVREEAYDLVQPKAMEAWKGVQFRELVEQEERITSVLSPEEIEACFDYNHHLKH 420

10 Query: 421 VDDIFERLGL 430
 VD IFERLGL
 Sbjct: 421 VDTIFERLGL 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 897> which encodes the amino acid sequence <SEQ ID 898>. Analysis of this protein sequence reveals the following:

15 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.3358(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 422/430 (98%), Positives = 428/430 (99%)

25 Query: 1 MIERYSRPMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
 M+ERYSRPEMAAIWTEENKY AWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL
 Sbjct: 1 MLERYSRPEMAAIWTEENKYHAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60

30 Query: 61 EIEQDTRHDVVAFTAVSETLGEERKQVWHYGLTSTDVVDYAGYLYKQANDIIRRDLENF 120
 EIEQDTRHDVVAFTAVSETLGEERKQVWHYGLTSTDVVDYAGYLYKQANDIIRRDLENF
 Sbjct: 61 EIEQDTRHDVVAFTAVSETLGEERKQVWHYGLTSTDVVDYAGYLYKQANDIIRRDLENF 120

35 Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPTTFGLKLATWYSEMKNRNERFEHAAAGVEAGKI 180
 TNIVADKA+EHK TIMMGRTHGVHAEPTTFGLKLATWYSEMKNRNERFEHAAAGVEAGKI
 Sbjct: 121 TNIVADKAREHKMTIMMGRTHGVHAEPTTFGLKLATWYSEMKNRNERFEHAAAGVEAGKI 180

40 Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRLDHAIFYFAVLASIATSIERMATEI 240
 SGAVGNFANIPPFVE+YVCDKLGIRPQEISTQVLPRLDHAIFYFAVLASIATSIERMATEI
 Sbjct: 181 SGAVGNFANIPPFVEEYVCDKLGIRPQEISTQVLPRLDHAIFYFAVLASIATSIERMATEI 240

45 Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVLWHE 300
 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENV+LWHE
 Sbjct: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVSLWHE 300

50 Query: 301 RDISHSSAERIITPDTTILIDYMLNRFNGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
 RDISHSSAERIITPDTTILIDYMLNRFNGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK
 Sbjct: 301 RDISHSSAERIITPDTTILIDYMLNRFNGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360

55 Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQBEIDELFNPIYYTKR 420
 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQBEIDELFNPIYYTKR
 Sbjct: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQBEIDELFNPIYYTKR 420

Query: 421 VDDIFERLGL 430
 VDDIF+RLG+
 Sbjct: 421 VDDIFKRLGI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-355-

Example 280

A DNA sequence (GBSx0306) was identified in *S.agalactiae* <SEQ ID 899> which encodes the amino acid sequence <SEQ ID 900>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -16.24    Transmembrane  145 - 161 ( 119 - 167)
    INTEGRAL    Likelihood = -9.98     Transmembrane  125 - 141 ( 119 - 144)
    INTEGRAL    Likelihood = -9.29     Transmembrane   28 - 44 ( 23 - 51)
    INTEGRAL    Likelihood = -7.01     Transmembrane  196 - 212 ( 193 - 220)
10  INTEGRAL    Likelihood = -6.21     Transmembrane   96 - 112 ( 88 - 116)
    INTEGRAL    Likelihood = -5.79     Transmembrane  249 - 265 ( 246 - 266)
    INTEGRAL    Likelihood = -2.87     Transmembrane  222 - 238 ( 222 - 238)
    INTEGRAL    Likelihood = -2.28     Transmembrane  279 - 295 ( 278 - 295)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.7496(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database:

    >GP:BAB13498 GB:AB028634 RNA polymerase [Flammulina velutipes]
      Identities = 83/336 (24%), Positives = 150/336 (43%), Gaps = 40/336 (11%)

25  Query: 152 ILLLI AFVSIGKNR-VYNFVQNLNYFEEVIWNYFEENPVKIKEKSLIK-----FLLTIS 205
      IL L  SI NR + ++ N ++ N+F+ + +K K L+I F++ +S
      Sbjct: 133 ILFLYLIYSILINRFILKWL DNSGIIYKININWFKNHMIKHINKMLVNIKFFNFIKLS 192

    Query: 206 FVFVIDFAMVRL-----LNFNIKSTILACSAILLAWLYQN-----KSVTEPFL 249
      + +I +++ L +NF+I+ I I ++ S+ F
30  Sbjct: 193 IITIIGISIMELFGIFGINFDIRIIIIINYLKTINSKGHLTIINMDQYSVLENSIHTIFY 252

    Query: 250 LKKLVIFVFIFFIATLIGNLKN-ELSILETPLLFISIFFTMDRIIALSKEMRDLI--ISKS 306
      + L+I+ IF L N+KN + +I +L+I IF I ++DL+ ++K
35  Sbjct: 253 INLLIIFLIFISLILYRNVKNIDTNIKRWIILYILIFLINIIFIFNHIYIKDLMNLNKY 312

    Query: 307 ILFFYDHNKIPKISILLSEIKEIKYLENVDIGE---LQLVRQMVIRLRLELEEFLLISDI 363
      IL Y D I S+ L ++K L+ ++I + V+ + I+ ++E L + I
40  Sbjct: 313 ILDYM LHIIVNSLFLFNKFDVK-LKRIN IYKSYSTVTVKDLEIKSKIEERSNELD IKLI 371

    Query: 364 YMKNG-YEKYIQFVQGNVYFINLE--LDKIPNYTNLKLILESIFD---HNNQKIFIPKL 416
      K G Y E YI ++ N+ ++ E L P Y N +E + + + F+ K+
45  Sbjct: 372 IAKYGSYENYINSIE-NINIVDEEFILKNYPEYINDSKFIEFLMELEPLFRDHTFVVKTI 430

    Query: 417 YEEYIYILISLGEVEKAKEIL---KEVSDYLTEESL 449
      YE L + K+IL KE+ DY+ + +L
      Sbjct: 431 YENLNSTNEKLEFL LANKDILSENKEIFDYVLQLNL 466

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 281

A DNA sequence (GBSx0308) was identified in *S.agalactiae* <SEQ ID 901> which encodes the amino acid sequence <SEQ ID 902>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>

```

-356-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 282

- 10 A DNA sequence (GBSx0309) was identified in *S.agalactiae* <SEQ ID 903> which encodes the amino acid sequence <SEQ ID 904>. This protein is predicted to be purK (purK). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9461> which encodes amino acid sequence <SEQ ID 9462> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:CAA04376 GB:AJ000883 purK [Lactococcus lactis]
 Identities = 208/347 (59%), Positives = 258/347 (73%), Gaps = 3/347 (0%)
 25 Query: 14 NSFKTIGIIGGGQLGQMMIAIAIYMGHKVITLDPASDCPASRVS-EVIVAPYDDVEALGT 72
 N+ +TIGIIGGGQLGQMMIAIAA YMGHKVITLDP +C A++VS E+IVAPYDDVE L
 Sbjct: 4 NTKQTIGIIGGGQLGQMMIAIAQYMGHKVITLDPNPNCSAAKVSDDELIVAPYDDVENLLR 63
 30 Query: 73 LAARCDVLTIEFENVADADGLDAVSAGQLPQGTDLRLISQNRIFEKDFLANKAGVTVAPY 132
 LA CDV+TYEFENV A L + ++PQG LL I+QNR FEK+FL N+A V VAP+
 Sbjct: 64 LAYACDVITYEFENVSAKALHEIEGCVRIPOGIRLLEITQNRFEKEFLTNEAKVNVAPW 123
 35 Query: 133 KVTSSLDLEGLDLTKTYVLKTATGGYDGHGQKVRSAEDLPEAQQLANSAQCVLEEFVN 192
 ++V S+ L +T+ VLKT TGGYDGHGQ V+ + E L A+ L ++CVLE+F++
 Sbjct: 124 QLVDSAEKLPET-VTRKQVLKTTTGGYDGHGQVVLNTDEKLSAAKSLTELESECVLEDFIS 182
 40 Query: 193 FDLEISVIVSGNGQDVTVPVQENIHRNLSKITIVPARISDQLADKAKEMAVQIAKKLQ 252
 F+ EISVI+SGNG + VFP+ EN HR NIL +TI PARIS ++ + A ++A IA+KL+
 Sbjct: 183 FERISVVISGNGHEYVVFPLAENEHRENILHQTISPARISAEITENAYKIATSIAEKLE 242
 45 Query: 253 LSGTLCVEMFATAD-DIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGVLGAPLPPIKLH 311
 LSG LCVEMF TAD I VNE+APRPHNSGH++IEACDF+QFD HI G+LG LP KL
 Sbjct: 243 LSGVLCVEMFLTADGQIYVNEIAPRPHNSGHFTIEACDFNQFDLHIKILGEDLPPEKLL 302
 Query: 312 AFAVMFNVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRMKGHVTV 358
 FA+M NVLGQHV+ ++ H H YGK +AKHNRMKGHVTV
 Sbjct: 303 KPAIMLNVLGQHVEAVKKLNHEHADWHQHDYKADAKHNRMKGHVTV 349

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 905> which encodes the amino acid sequence <SEQ ID 906>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----

-357-

bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 344/369 (93%), Positives = 353/369 (95%)

Query: 1 MRNKEKSQRSQAMNSFKTIGIIGGGQLGQMMATAIYMGHKVITLDPASDCPASRVSEVI 60
 MRNKEKSQRSQ +NSFKTIGIIGGGQLGQMMATAIYMGHKVITLDPASD PASRVSEVI
 10 Sbjct: 1 MRNKEKSQRSQVNSFKTIGIIGGGQLGQMMATAIYMGHKVITLDPASDSPASRVSEVI 60

Query: 61 VAPYDDVEALGTLAARCDVLTVEFENVADGLDAVVSAGQLPQGTDLIRISQNRIFEKDF 120
 VAPYDDVEALG LAARCDVLTVEFENVADGLDAVSA QLPQGTDLIRISQNRIFEKDF
 15 Sbjct: 61 VAPYDDVEALGQLAARCDVLTVEFENVADGLDAVVSACQLPQGTDLIRISQNRIFEKDF 120

Query: 121 LANKAGVTVAPYKVTSSLDLEGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLA 180
 LANKAGVTVAPYKVTSSLDL GLDLTKTYVLKT TGGYDGHGQK+IRSAEDLPEAQQLA
 Sbjct: 121 LANKAGVTVAPYKVTSSLDLGGDLTKTYVLKTETGGYDGHGQKIIRSAEDLPEAQQLA 180

Query: 181 NSAQCVLEEFVNFDEISVIVSGNGQDVTVPVQENIHRNNILSKTIVPARISDQLADKA 240
 NSAQCVLEEFVNFDEISVIVSGNG+DVTVPVQENIHRNNILSKTIVPARISDQLADKA
 20 Sbjct: 181 NSAQCVLEEFVNFDEISVIVSGNGKQDVTVPVQENIHRNNILSKTIVPARISDQLADKA 240

Query: 241 KEMAVQIAKKLQLSGTLCVEMFATADDIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGV 300
 K+ AVQIAKKLQLSGTLCVEMF TADDIIVNEIAPRPHNSG YSIEACDFSQFDTHILGV
 25 Sbjct: 241 KKTAVQIAKKLQLSGTLCVEMFTTADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300

Query: 301 LGAPLEPPIKLHAPAVMFENVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRKMGHVTVFS 360
 LGAPLE I+LHAPAVM NVLGQHVQQA D+VA+NPSAHLHMYGKLEAKHNRKMGHVTVF+
 30 Sbjct: 301 LGAPLEQIQLHAPAVMLNVLGQHVQQAIDYVAKNPSAHLHMYGKLEAKHNRKMGHVTVFA 360

Query: 361 DVPDEVEEF 369
 DEV+EF
 35 Sbjct: 361 KDADEVKEF 369

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 283

A DNA sequence (GBSx0310) was identified in *S.agalactiae* <SEQ ID 907> which encodes the amino acid
 40 sequence <SEQ ID 908>. This protein is predicted to be phosphoribosylaminoimidazole carboxylase
 catalytic subunit (purE). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3572 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12462 GB:Z99107 phosphoribosylaminoimidazole carboxylase I
 [Bacillus subtilis]
 Identities = 106/162 (65%), Positives = 128/162 (78%)

55 Query: 33 MQPIISIIMGSKSDWTTMQKTAEVLNFGIAYEKKVSAHRTPDLMFKHAEAEARGGIKI 92
 MQP++ IIMGS SDW TM+ ++LD + YEKKVSAHRTPD MF++AE AR RGIK+
 Sbjct: 1 MQPLVGIIMGSTSDWETMKHACDILDELNVPEYKVVSAHRTPDFMFYAETARERGIKV 60

Query: 93 ILAGAGGAHLPGMVAAKTTLPVIGVPVKSRALSGLDLSIVQMPGGVPVATMAIGEAG 152
 60 ILAGAGGAHLPGM AAKTTLPVIGVPV+S+AL+G+DSL SIVQMPGGVPVAT +IG+AG

-358-

Sbjct: 61 IIAGAGGAAHLPGMTAAKTTLPVIGVPVQSKALNGMDSLISIVQMPGGVPVATTSIGKAG 120

Query: 153 ATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194

A NA L A +ILS D++LA L E + ESS++L+

Sbjct: 121 AVNAGLLAAQILSAFDEDLARKLDERRENTKQTVLESSDQLV 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 909> which encodes the amino acid sequence <SEQ ID 910>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.08 Transmembrane 36 - 52 (34 - 52)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2232 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA04375 GB:AJ000883 purE [Lactococcus lactis]
 Identities = 105/158 (66%), Positives = 131/158 (82%)
 Query: 46 ISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVVSARHTPDLMPKHAEEARGGIKIIAG 105
 ++IIMG SDWATM++TA++LD+FG+AYEKKVVSARHTP LM + + +AR RG K+IIAG
 Sbjct: 4 VAIIMGSSDWATMKETAKILDDFGLAYEKKVVSARHTPALMAEFSSQARERGYKVIIAG 63
 Query: 106 AGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVATMAIGEAGATNA 165
 AGGAAHLPGMV+A+T +PVIGVP+KSRLSGLDSLISIVQMP GVPVATMAIGEAGA NA
 Sbjct: 64 AGGAAHLPGMVSAQTLVPVIGVPIKSRLSGLDSLISIVQMPAGVPVATMAIGEAGAKNA 123
 Query: 166 ALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203
 AL AL++L+ ++NL L + ++ EES+ L+
 Sbjct: 124 ALFALQLLANTNENLIQKLLVYRAAAQEMVEESNKALL 161

An alignment of the GAS and GBS proteins is shown below:

Identities = 162/169 (95%), Positives = 164/169 (96%), Gaps = 1/169 (0%)
 Query: 27 PLYLNIMQ-PIISIIMGSKSDWTTMQKTAEVLDNFGIAYEKKVVSARHTPDLMPKHAEEA 85
 PL + IM+ PIISIIMGSKSDW TMQKTAEVLDNFGIAYEKKVVSARHTPDLMPKHAEEA
 Sbjct: 35 PLCILIMKTPIIISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVVSARHTPDLMPKHAEEA 94
 Query: 86 RGRGIKIIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT 145
 RGRGIKIIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT
 Sbjct: 95 RGRGIKIIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRLSGLDSLISIVQMPGGVPVAT 154
 Query: 146 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194
 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESS ELI
 Sbjct: 155 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 284

A DNA sequence (GBSx0311) was identified in *S.agalactiae* <SEQ ID 911> which encodes the amino acid sequence <SEQ ID 912>. This protein is predicted to be phosphoribosylglycinamide synthetase (purD). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-359-

bacterial cytoplasm --- Certainty=0.1966(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 298/419 (70%), Gaps = 7/419 (1%)

10 Query: 1 MKLLVVGSGGREHAIAKKLLASKVDQVFVAPGNDGMTLDGLDLVNIGISEHSRLIDFVK 60
 MK+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ F +
 Sbjct: 1 MKILVIGSGGREHALAKKFMESEPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

15 Query: 61 ENEIAWTLIGPDDALAAGIVDGFNSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
 I T +GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

20 Query: 121 YCTFSDFEKAKAYIEEQAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180
 Y TF E A AY++E+G P+V+KADGLA GKGVA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

25 Query: 181 ARVVIEEFLDGEEFSLFAPFANGDKFYIMPTAQDHKRAYDGDGKGLNTGGMGAYAPVPHLPQ 240
 +VVIEEFLDGEEFSLF+ + K Y MP AQDHKRA+D DKG NTGGMGAY+PV H+ +
 Sbjct: 176 GKVVIEEFLDGEEFSLFSFIHDGKIYPMPIAQDHKRAFDGDKGPNNTGGMGAYSFVLHISK 235

30 Query: 241 SVVDTAETIVKPVLEGMIAGRPYLGVLVYAGLILITADGPKVIEFNSRFGDPETQIILPR 300
 VV+ A+E +VKP + GMI EG+ + GVLVYAGLILT DG K IEFN+RFGDPETQ++LPR
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILITEDGVKTIEFNARFGDPETQVVLPR 295

35 Query: 301 LTSDFQAQNIIDIMGIEFYITWQKDGVTLGVVVASEGYPLDYKGVPLPEKTDGDIITYY 360
 L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP + G+ LPE +G + YY
 Sbjct: 296 LKSDLAQAIIDILAGNEFTLEWLESQVTLGVVVAAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDITGLFYRNDIGSKAI 419
 AG EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI
 Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 913> which encodes the amino acid sequence <SEQ ID 914>. Analysis of this protein sequence reveals the following:

40 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 5 - 21 (5 - 21)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

55 Query: 50 LKLLVVGSGGREHAIAKKLLASKGVVDQVFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+
 Sbjct: 1 MKILVIGSGGREHALAKKFMESEPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
 I F+GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

60 Query: 170 YCTFSDFEKAKAYIEEQAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
 Y TF E A AY++E+G P+V+KADGLA GKGVA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGEEFSLFAPFANGDKFYIMPTAQDHKRAFDGDKGPNNTGGMGAYAPVPHLPQ 289

-360-

+VVIEEFLDGEFSLF+F + K Y MP AQDHKRAFD DKGENTGGMGAY+VH +
 Sbjct: 176 GKVVIEEFLDGEFSLFSLFSFIHDGKIYPMPAQDHKRAFDDEKENTGGMGAYSPVLHISK 235
 Query: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNSRFGDPETQIILPR 349
 5 VV+ A+E +V+P + GM+ EG+ + GVLV GLILT DG K IEFN+RFGDPETQ++LPR
 Sbjct: 236 EVVNEALEKVVKFTVAGMIEEGKSFTIGVLYAGLILTEDGVKTIEFNARFGDPETQVVLP 295
 Query: 350 LTSDFQNIIDDIMMIEPYITWQKDGVTGLGVVVASEGYPFDYKGVPLPEKTDGDIITYY 409
 10 L SD AQ I DI+ G EP + W + GVTGLGVVVA+EGYP + G+ LPE +G + YY
 Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTGLGVVVAEGYPSQAKLGLILPEIPEG-LNVYY 354
 Query: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDITGLFYRNDIGSKAI 468
 AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI
 15 Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

An alignment of the GAS and GBS proteins is shown below:

Identities = 399/421 (94%), Positives = 408/421 (96%)

Query: 1 MKLLVVGSGGREHAIKLLASKDQVDFVAPGNDGMTLDGLDLVNIIGISEHSRLIDFVK 60
 20 +KLLVVGSGGREHAIKLLASK VDQVDFVAPGNDGMTLDGLDLVNI +SEHSRLI F K
 Sbjct: 50 LKLLVVGSGGREHAIKLLASKGVDQVDFVAPGNDGMTLDGLDLVNIIVVSEHSRLIAFAK 109
 Query: 61 ENEIAWTILIGPDDALAAGIVDGFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
 25 ENEI+W IGPDDALAAGIVD FNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA
 Sbjct: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
 Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180
 30 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG
 Sbjct: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
 Query: 181 ARVVIEEFLDGEFSLFAFANGDKFYIMPTAQDHKRAYDGDGKGLNTGGMGAYAPVPHLPQ 240
 ARVVIEEFLDGEFSLFAFANGDKFYIMPTAQDHKRA+DGDGK NTGGMGAYAPVPHLPQ
 35 Sbjct: 230 ARVVIEEFLDGEFSLFAFANGDKFYIMPTAQDHKRAFDGDKGPNITGGMGAYAPVPHLPQ 289
 Query: 241 SVVDTAVETIVKPVLEGMIAEGRPYLGVLYAGLILTADGPKVIEFNSRFGDPETQIILPR 300
 SVVDTAVE IV+PVLEGM+AEGRPYLGVLY GLILTADGPKVIEFNSRFGDPETQIILPR
 40 Sbjct: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNSRFGDPETQIILPR 349
 Query: 301 LTSDFQNIIDDIMMIEPYITWQKDGVTGLGVVVASEGYPFDYKGVPLPEKTDGDIITYY 360
 40 LTSDFQNIIDDIMMIEPYITWQKDGVTGLGVVVASEGYP DYKGVPLPEKTDGDIITYY
 Sbjct: 350 LTSDFQNIIDDIMMIEPYITWQKDGVTGLGVVVASEGYPFDYKGVPLPEKTDGDIITYY 409
 Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDITGLFYRNDIGSKAIKE 421
 AG KF+ENS+ LLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDITGLFYRNDIGSKAI+E
 45 Sbjct: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDITGLFYRNDIGSKAIRE 470

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 285

50 A DNA sequence (GBSx0312) was identified in *S.agalactiae* <SEQ ID 915> which encodes the amino acid sequence <SEQ ID 916>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 55 INTEGRAL Likelihood = -1.28 Transmembrane 235 - 251 (235 - 251)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-361-

>GP:AAA23257 GB:M81878 unknown [Clostridium perfringens]

Identities = 66/258 (25%), Positives = 119/258 (45%), Gaps = 9/258 (3%)

5 Query: 1 MTIYDQIESALDLMTDLEREIACYFMGQPIISKDALASTIVTKQLHISQAALTRFAKKCGF 60
 M I +Q+E+ T E+ + Y + + +I+ K+ + +A +TRF KK GF
 Sbjct: 1 MGILEQLENPKFKATKSEKTLIEYIKSDLDNIIYKSISIIAKBSGVGEATTTRFTKKLGF 60

10 Query: 61 KGYREFVFEYLKS-HETISQQLYGLQNDNTKKVFMNYQEMISKSADI-----IDEEQL 112
 G+++F K + + L + V +M+ S +I ID + +
 Sbjct: 61 NGFQDFKVTLAKEISNKKNTSIIINLHVHRDESVTETANKMLKSSINILEQTVKQIDLDLM 120

15 Query: 113 LEVSHMIEQADRVYFYGKGSSSLVAKEFKIRLMRLGVCEALDDTDSFSWTNSIVNDRCL 172
 + +I A RVYF G G S + A + + MR+G + D+ + +SI ND +
 Sbjct: 121 CKCRDLIMNAKRVIYFIGYSGIAATDINYKFMRI GFTTVPVTD SHTMVMSSITNDDDV 180

20 Query: 173 VIAFSLSGNTNSVIGALKIASCHGAKTVLFTK-QPHTIDYAFDKIIQVASARHLDYGNRI 231
 ++A S SG T VI +K A +G K + T+ ++ D + SA + I
 Sbjct: 181 IVAISNSGTTKEVIKTVKQAKENGTKIITLTEDSDNPLRKLSDYELTYTSAETIFETGSI 240

25 Query: 232 SPQIPMLIMVDIIYAQFL 249
 S +IP + ++D++Y + +
 Sbjct: 241 SSKIPQIFLLDLLYTEVI 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 917> which encodes the amino acid
 25 sequence <SEQ ID 918>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.88 Transmembrane 243 - 259 (242 - 261)

30 ----- Final Results -----

bacterial membrane --- Certainty=0.2954(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 A related sequence was also identified <SEQ ID 9093> which encodes the amino acid sequence <SEQ ID
 9094>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 56

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.88 Transmembrane 239 - 255 (238 - 257)

40 ----- Final Results -----

bacterial membrane --- Certainty= 0.295(Affirmative) < succ>

bacterial outside --- Certainty= 0.000(Not Clear) < succ>

bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below:

Identities = 138/263 (52%), Positives = 189/263 (71%), Gaps = 2/263 (0%)

50 Query: 6 QIESALDLMTDLEREIACYFMGQPIISKDALASTIVTKQLHISQAALTRFAKKCGFKGYRE 65
 +IE++L+ MT LE+ IA +F+ ++ L ++ + K+LHISQAALTRFAKKCGF GYR
 Sbjct: 14 KIEASLEHMTSLEKGTAFHFFITTLTPQELTASEIVKRLHISQAALTRFAKKCGFTGYRA 73

55 Query: 66 FVFEYLKSHETISQQLYGLQNDNTKKVFMNYQEMISKSADIIDEEQLLEVSHMIEQADRV 125
 F F+YL S + + + TK+V M+Y +I+K+ +++EE+LL ++ +I+ ++RV
 Sbjct: 74 FAFDYLHSLQESQETFQSIHLELTKRVLMDYDALINKTVELVNEEKLNLAKLIDSSERV 133

60 Query: 126 YFYGKGSSSLVAKEFKIRLMRLGVCEALDDTDSFSWTNSIVNDRCLVIAFSLSGNTNSV 185
 YF+GKGSS LVA+E K+R MRLG+IC+A DTD F+W NS+VN+ CLV FSLSG TNSV
 Sbjct: 134 YFFGKGSSGLVAREMKLRFMRLGLICDAYSDDTGTWANSLVNENCLVFGFSLSGKTNSV 193

Query: 186 IGALKIASCHGAKTVLFTKQPHT-IDYAFDKIIQVASARHLDYGNRISQIPMLIMVDII 244
 I AL AS GAKTVL T T D + D II V+S L YGNR+SPQ P+LIM+DII
 Sbjct: 194 ITALHQASQRGAKTVLLTTDNQTEFDDSLD-IIPVSSTHQLHYGNRVSPQFPLIMMDII 252

-362-

Query: 245 YAQFLDINKIEKERIFRETIQR 267
 YA L I+K KE+IF+ TII +
 Sbjct: 253 YAYVLAIIDKPKHKEKIFKNTIIDK 275

SEQ ID 916 (GBS320) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 5; MW 33kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 7; MW 58kDa) and in Figure 160 (lane 7 & 8; MW 58kDa).

GBS320-GST was purified as shown in Figure 224, lane 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 286

A DNA sequence (GBSx0313) was identified in *S.agalactiae* <SEQ ID 919> which encodes the amino acid sequence <SEQ ID 920>. This protein is predicted to be xylan esterase 1 (cephalosporin-C). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB68821 GB:AF001926 xylan esterase 1 [Thermoanaerobacterium sp.
 'JW/SL YS485']
 Identities = 133/299 (44%), Positives = 188/299 (62%), Gaps = 1/299 (0%)

Query: 5 MSLDDMREYLGQDQIPEDFDDFWKKQTMKYQG-NIEYRLDKKDFNITFAQAYDLHFKGSN 63
 M L +REY G + PEDFD++W + + + + L + F ++FA+ YDL+F G
 Sbjct: 6 MPLQLREYTGTPNCPEDFDEYWNRALDEMRSVDPKIELKESSFQVSFAECYDLYFTGVR 65

Query: 64 NSIVYAKCLFPKINKPYPVVFYFHGYQNQSPDWSQNLNYVAAGYGVVSMQVVRGQAGQSQD 123
 + ++AK + PKT +P + FHGY + S DW+D+LNYVAAG+ VV+MDVRGQ GQSQD
 Sbjct: 66 GARIHAKYIKPKTEGKHPALIRFHGYSSNSGDNWDLKLYVAAGFTVVAMQVVRGQGGQSQD 125

Query: 124 KGHFDGITVKQIVRGMISGPNHLFYKDIYLDVFLIDIIATLESVDSNQLYSYGWSQGG 183
 G G T+ G I+RG+ + + + I+LD QL I+ + VD + + + G SQGG
 Sbjct: 126 VGGVTGNTLNHIIIRGLDDADNMLFRHIFLDTAQLAGIVMMPEVDEDRVGMGPGSQGG 185

Query: 184 ALALIAAALNPKIVKTVAVYPFLSDFRRLVLDLGGVSEPYDELFRYFKYSDPFHKTENNVL 243
 L+L AAL P++ K V+ YPFLSD++RV DL Y E+ YF+ DP H+ EN V
 Sbjct: 186 GLSLACAALEPRVRKVVSEYPFLSDYKRVWDLDLAKNAYQEITDYFRLFDPRHERENEVF 245

Query: 244 KTLAYIDVKNFARISCPVLLTALKDDICPPSTQFAIFNRLTSTKKHLLLPDYGHDP 302
 L YIDVKN A RI V++ L D +CPPST FA +N + S K + PDYGH+PM
 Sbjct: 246 TKLG YIDVKNLAKRIKGDVLMCVGLMDQVCPSTVFAAYNNIQSKKDIKVYPDYGHEPM 304

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-363-

Example 287

A DNA sequence (GBSx0314) was identified in *S.agalactiae* <SEQ ID 921> which encodes the amino acid sequence <SEQ ID 922>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -5.73    Transmembrane 128 - 144 ( 126 - 145)

----- Final Results -----
10      bacterial membrane --- Certainty=0.3293 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAA23256 GB:M81878 unknown [Clostridium perfringens]
    Identities = 78/160 (48%), Positives = 110/160 (68%)

Query: 131 CLTIGTGIGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIADVAKAHGDEI 190
      CLTIGTGIGG LIID V HGFNSA E+GYM ++ + QD+AS +AL+ +VA G E
20  Sbjct: 18 CLTIGTGIGGALIIDGKVLHGFSNSAGEIGYMMVNGENIQDIASASALVKINVALRKGVEP 77

Query: 191 SRWDGRRIFQEAKKGNEKCIASIDRMINYLGQGIANMVVYVNPKEKVVLGCGGIMAQKDYDYLQ 250
      S DGR + + G+ C +++++ + L GI+N+VY++NPE VVLGGGIMA+++ +
25  Sbjct: 78 SSIDGRYVLDNYENGDLICKEEVEKLADNLALGISNIVYLINPEVVVLGGGIMAREEVFR 137

Query: 251 DKLSESLKRNLVTSLEAKTAIVFAQHENQAGMLGAYYHFK 290
      + SL++ L+ S+ T I FA+ +N AGM GAYY+FK
30  Sbjct: 138 PLIENSLRKYLIESVYNNTKIAFAKLKNTAGMKGAYYNFK 177

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 923> which encodes the amino acid sequence <SEQ ID 924>. Analysis of this protein sequence reveals the following:

```

Possible site: 22

>>> Seems to have no N-terminal signal sequence
35  INTEGRAL    Likelihood = -4.30    Transmembrane 128 - 144 ( 127 - 145)
    INTEGRAL    Likelihood = -0.11    Transmembrane 227 - 243 ( 227 - 243)

----- Final Results -----
40      bacterial membrane --- Certainty=0.2720 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

45  >GP:BAB04516 GB:AP001509 glucose kinase [Bacillus halodurans]
    Identities = 97/291 (33%), Positives = 155/291 (52%), Gaps = 14/291 (4%)

Query: 5 LAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLAGVAI 64
      + ID+GGT IK L+S+ G+++ +E TEA +G ++ K+ L + D AG+ I
50  Sbjct: 3 VGIDLGGTKIKAALVSDAGEIISVQECPTAAQGFEEVMNKMMSLTEKVTDRHQPFGAGIGI 62

Query: 65 SSAGMVNPDGEIFYAGPQIPNYAGTQPKKEIBETFGLPCEVENDVNCAGLAEAISGSAS 124
      + G ++ EG I + P +P + +E F P +++ND N A LAEA+ GS +
55  Sbjct: 63 GAPGLSSTEGTLL-SPPNLPGWHDHILVDRFQEQFQCPVKLDNDANVAALAEALLGSGQ 121

Query: 125 DYPVALCLTIGTGIGGCLLFNSQVFHGSSSHSAACEVG-----YLHLSDGQFQDLAS 174
      + LTI TGIGG + + + HG+S A E+G + +L+ G + LAS
60  Sbjct: 122 GFTSVFYLTISTGIGGGYVLDGSIHVGASDYAGEIGNMIVQPNGYQHANLNPGLSLEGLAS 181

Query: 175 TTALVQEVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISQVDYLGQGIANICYVNP 234
      TA+ + +G + R +F+Q + GD + + +DYL GIANI + +NP+
60  Sbjct: 182 GTAIGRMARERFG---VEGGTREVFDQIRRGDHDQMRLVEEAMDYLAIGIANIAHTINPD 238

```

-364-

Query: 235 VVVLGGGIMAQKDYLDKLTALDSYLVSSLAKKTQLKFFASHGNNAGILGA 285
 V VLGGG+M D + +K + YL LA+ T + A G ++G+LGA
 Sbjct: 239 VFVLGGGVMNADDLILPIVKEKVSRYLPGLAQSTTIVKAKLGGDSGVVLA 289

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 192/292 (65%), Positives = 237/292 (80%)

Query: 1 MTRTV AIDIGGTMIKHGIVDNLGCI VEASELATEAYKGGPGILQKVCQIIDNYLAEGSID 60
 M +AIDIGGT IK+G++ G ++E E+ATEAYKGGP IL+KV ++ Y + +
 10 Sbjct: 1 MKHYLAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLA 60

Query: 61 GIAISSAGMVPDEGCIFYSGPQIPNYAGTQFKVLEDYQVRTEIENDVNCAGLAEAVS 120
 G+AISSAGMV+EDEG IFY+GPQIPNYAGTQFKK +E+T+ + E+ENDVNCAGLAEA+S
 15 Sbjct: 61 GVAISSAGMVNPDEGEIFYAGPQIPNYAGTQFKKEIEETFGLEPCEVENDVNCAGLAEAIS 120

Query: 121 GSAKDESSIALCLTIGTGIGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIA 180
 GSAKD +ALCLTIGTGIGGCL+ + VFHG S+SACEVGY+HLSHG FQDLASTTAL+
 Sbjct: 121 GSAKDYPVALCLTIGTGIGGCLLFNSQVFGSSSHSACEVGYLHLSHGQFQDLASTTALVQ 180

20 Query: 181 DVAKAHGDEISRWDGRRIFQEA KKGNEKCIASIDRMINYLGGQIANMVYVNPKEKVVVLGG 240
 +V A+GD+IS+WDGRRIF++AK G+ CIA+I + ++YLGGQIAN+ YVNP VVLGG
 Sbjct: 181 EVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISKQVDYLGQGIANICYVVPNPVNVVLGG 240

25 Query: 241 GIMAQKDYLDKLSLSESLKRNLTSLAEKTAIVFAQHENQAGMLGAYYHFKNR 292
 GIMAQKDYLD KL +L LV+SLA+KT + FA H N AG+LGAYYHFK +
 Sbjct: 241 GIMAQKDYLDKLTALDSYLVSSLAKKTQLKFFASHGNNAGILGAYYHFKQK 292

SEQ ID 922 (GBS331) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 2; MW 35.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 3; MW 61kDa).

The GBS331-GST fusion product was purified (Figure 209, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 309), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 288

A DNA sequence (GBSx0315) was identified in *S.agalactiae* <SEQ ID 925> which encodes the amino acid sequence <SEQ ID 926>. This protein is predicted to be a acylneuraminate lyase (nanA). Analysis of this protein sequence reveals the following:

40 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.0894(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA69950 GB:Y08695 putative acylneuraminate lyase [Clostridium
 tertium]
 50 Identities = 162/225 (72%), Positives = 191/225 (84%)

Query: 1 MKDLQKYQGIIIPAFYACYDDKGDICPERVKALTNFYIDKGVQGLYVNGSSGECIYQSVAD 60
 M++L+KY+GIIPAFYACYDD+G I PER + T Y IDKGV+GLYV GSSGECIYQS +
 55 Sbjct: 1 MRNLEKYKGIIIPAFYACYDDGKISPRTQMFTQYLIDKGVKGLYVCGSSGECIYQSKEE 60

-365-

Query: 61 RKLVLNVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120
 RK+ LENVM VAKGK+T+IAHV CNNT+DS ELA HAE+IGVDAIA+IPPIYF LP+Y+I
 Sbjct: 61 RKITLENVMKVAKGKITIIAHVGCNNTDSEELAEHAEISIGVDAIASIPPIYFHLDPYSI 120

Query: 121 ADYWNTISQAAPQTDFTIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180
 A+YWN IS AAP TDFIYNIPQLAGV L +LY++ML+NP+VIGVKNSSMPVQDIQ F
 Sbjct: 121 AEYWNDISNAAPNTDFTIYNIPQLAGVGLGINLYKQMLKNPRVIGVKNSSMPVQDIQMFK 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAAGIGGTYGVMPELYLTINQ 225
 I G+ +VFNGPDEQF+ GR+MCA GIGGTY VMPEL+L ++
 Sbjct: 181 DISGDESUVFNGPDEQFVAGRIMGADGGIGGTYAVMPELFLAADK 225

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 927> which encodes the amino acid
 sequence <SEQ ID 928>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 238/304 (78%), Positives = 263/304 (86%)

Query: 1 MKDLQKYQGIIPAFYACYDDKGDICPERVKALINYFIDKGVQGLYVNGSSGECIYQSVAD 60
 M DL KYQGIIPAFYACYDD+G+I PERV+ALT Y+IDKGVQGLY+NGSSGECIYQSV D
 Sbjct: 1 MIDLTQKYQGIIPAFYACYDDQGNISPERVRALTQYYIDKGVQGLYINGSSGECIYQSVFD 60

Query: 61 RKLVLNVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120
 R+LVLENVM+VAKGKLT+I HVACNNTKDS+ELA H+E +GVDAIAAIPPIYFRLPEYA+
 Sbjct: 61 RQLVLNVMAVAKGKLTIIHVACNNTKDSIELAAHSERLGVDAIAAIPPIYFRLPEYAV 120

Query: 121 ADYWNTISQAAPQTDFTIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180
 ADYWN IS AAP TDFIYNIPQLAGVALT LY+ ML N +VIGVKNSSMPVQDIQ F
 Sbjct: 121 ADYWNAISSAAPHTDFTIYNIPQLAGVALTPSLYKTM LANKRVIGVKNSSMPVQDIQTFC 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAAGIGGTYGVMPELYLTINQLIVDKDLEKARELQF 240
 AIGG++HIVFNGPDEQFLGGRLMGAAAAGIGGTYG MP EL+L LNQLI DKDLEKA+ LQ+
 Sbjct: 181 AIGGDDHIVFNGPDEQFLGGRLMGAAAAGIGGTYGAMPELFLRLNQLIADKDLEKAKALQY 240

Query: 241 TINDIITKLCSGHGNMYAVIKAVLEINEQLTIGSVRLPLASVTEEDKPIIKEAAEMIRHA 300
 TIN+II L S HGNY VIK VL INE L IGSVR PLA + EED+ I + AA +I A
 Sbjct: 241 TINEIIGVLVSAHGNMYGVIKEVLRINEGLDIGSVRSPLAELVEEDRVICQRAAALINQA 300

Query: 301 KKQF 304
 K+ F
 Sbjct: 301 KETF 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 289

A DNA sequence (GBSx0317) was identified in *S.agalactiae* <SEQ ID 929> which encodes the amino acid
 sequence <SEQ ID 930>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98 (79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40 (21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196 (172 - 200)

-366-

INTEGRAL Likelihood = -5.10 Transmembrane 160 - 176 (158 - 179)
 INTEGRAL Likelihood = -4.35 Transmembrane 110 - 126 (106 - 130)

----- Final Results -----

5 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:BAB05827 GB:AP001514 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 40/148 (27%), Positives = 74/148 (49%), Gaps = 4/148 (2%)

15 Query: 14 VNNPFMQGCVVFDLALLNLLFMI-TCLPLVTIG--AAKISLYRTLWQKLEGD-QINLLI 69
 +++ F Q C+ ++ LA +NLL++ T L LV +G A +++ L + G+ +
 Sbjct: 6 MSSRFYQTCDWIWKLAYINLLWLSGTLGLVVLGFLPATTAMFTVLRKWNFTGNPDVAITR 65

Query: 70 LYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTS 129
 + + K E+ + LLG V L ++ F+ L G + L + YAFLL++T
 20 Sbjct: 66 TFFQAYKNEFLKINLLGAVLLLGAYILYFNMYLGTVEGTVMVLSLGGWYAFLLIYIITL 125

Query: 130 IYLFPMARYEMSLDVTVKKSFIACLN 157
 Y+ P Y + L +K + I+ +N
 25 Sbjct: 126 FYIIPAYVHYNLKLQYIKTALIIGFVN 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 931> which encodes the amino acid sequence <SEQ ID 932>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -14.86 Transmembrane 117 - 133 (108 - 139)
 INTEGRAL Likelihood = -7.48 Transmembrane 30 - 46 (21 - 54)
 INTEGRAL Likelihood = -6.90 Transmembrane 88 - 104 (83 - 105)
 INTEGRAL Likelihood = -6.26 Transmembrane 165 - 181 (151 - 187)
 35 INTEGRAL Likelihood = -5.89 Transmembrane 189 - 205 (182 - 207)

----- Final Results -----

35 bacterial membrane --- Certainty=0.6944(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB05582 GB:AP001513 unknown conserved protein in bacilli
 [Bacillus halodurans]
 Identities = 59/194 (30%), Positives = 93/194 (47%), Gaps = 11/194 (5%)

45 Query: 17 SKWMRASAAFLDILVFNLLFVL-SCLPLLTIGV--AKMALYASLLDWREGQVS-QLVTTY 72
 +K M+ + L+ NLL++L S + + +GV A +L+A W + + L TY
 Sbjct: 8 TKIMKLFWEIMRLVYLNLLWLLFSFIGGIILGMPATASLFAVFRKWYQKEDDFPLFQTY 67

50 Query: 73 SSHFKYFYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFQGFVKVLCVAVLFLVVILFLY 132
 + FK FK +GL + I I LD+ L+ S + Q + A+ F+ ++ LY
 Sbjct: 68 LNEFKRSPKIANLVGLTLVLIGGILYLDVLLLLGTSHWIGQLLLMGVGALSFIYLVTLTY 127

55 Query: 133 AYPQAVKRDLSLSTLFKRSFLLAGLFFPWSEAFALFICLTIFSLQL----SLLTLFGGVS 188
 +P V DLS FK SFLG+ P+ L I L++ +L LL LF S
 Sbjct: 128 IFTPLVHFDLSYKQYFKHSFLL-GVLQPFRTLLMLTSLSALLFLTFPILPLF-AAS 184

Query: 189 LLAIIGISSLTLY 202
 +A + + S + Y
 60 Sbjct: 185 FMAALTMWSFLFGY 198

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 117/210 (55%)

-367-

Query: 3 KANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMITCLPLVTIGAAKISLYRTLWQKLEG 62
 K L+ ++F +++ +M+ +FDL + NLLF+++CLPL+TIG AK++LY +L EG
 Sbjct: 4 KKQGLLHSLFKLDSKWMRASAAFLDLLVFNLLFVLSCLPLLTIGVAKMALYASLLDWREG 63

Query: 63 DQTNLLILYIKHLKKEWFQGMMLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFL 122
 + L+ Y H K + G+ LGL+EL I+ + + DL ++ Q G + K+ C A L
 Sbjct: 64 QVSQVLVTYSSHFYFKYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFOGFKVLCVAVL 123

Query: 123 LLTVMTSIYLFPMARYEMSLLDTVKKSFIMACLNKWTGVLMLLIMTWFMVQSSLLF 182
 L V+ +Y +P A + ++SL K+SF++A L W+ + + +T F + S L
 Sbjct: 124 FLVVILFLYAYPOAVKRDLSLSTLFKRSFLLAGLFFPWSFAFLAFICLTIFSLQLSLLTL 183

Query: 183 MLTVSAIFIFAYTAFAYFKIIILQKQFAYF 212
 VS + I ++ Y +II++ F
 Sbjct: 184 FGGVSLLAIGISSLTLYLYIIMESLLRRF 213

A related GBS gene <SEQ ID 8535> and protein <SEQ ID 8536> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 McG: Discrim Score: 3.27
 GvH: Signal Score (-7.5): -4.23
 Possible site: 46
 >>> Seems to have an uncleavable N-term signal seq

ALOM program count: 5 value: -9.45 threshold: 0.0

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98 (79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40 (21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196 (172 - 200)
INTEGRAL	Likelihood = -5.10	Transmembrane	160 - 176 (158 - 179)
INTEGRAL	Likelihood = -4.35	Transmembrane	110 - 126 (106 - 130)
PERIPHERAL	Likelihood = 5.89		142

modified ALOM score: 2.39

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.4779(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

ORF00072(364 - 828 of 1260)
 EGAD|108353|BS3003(14 - 171 of 222) hypothetical protein {Bacillus subtilis}
 OMNI|NT01BS3507 conserved hypothetical protein GP|2635493|emb|CAB14987.1||Z99119 similar to
 hypothetical proteins from B. subtilis {Bacillus subtilis}
 GP|2293197|gb|AAC00275.1||AF008220 YteU {Bacillus subtilis} PIR|D69991|D69991 conserved
 hypothetical protein yteU - Bacillus subtilis
 %Match = 5.9
 %Identity = 26.6 %Similarity = 50.6
 Matches = 42 Mismatches = 74 Conservative Sub.s = 38

270	300	330	360	390	417	441	471
IMSKKGY*KC*WRKKYREYIVK	KANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMI	-TCLPLVTIG--AAKISLYRTLW					
			: : :				: : :
			MEHDGSLGRMLRFCEWIMRFAYTNLLWLFFTLGLGVFGIMPATAALFAVMR				
			10 20 30 40 50				

498	528	558	588	618	648	678	708
QKLEG-DQTNLLILYIKHLKKEWFQGMMLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTSIYLFPM							
: : : : : : : : : : : : : :							
KWIQGDQNVFLKTFVQYKGEFFRSNLLGAVLALIGVYIIDLALI-YPSHFLHLIRFAIMIFGFLFVSMFLFYVFP							
	70	80	90	100	110	120	130

738	768	798	828	858	888	918	948
ARYEMSLLDTVKKSFIMACLNKWTGVLMLLIMTWFMVQSSLLFMLTVSAIFIFAYTAFAYFKIIILQKQFAYFSKQ							

-368-

```

      ::      || |:::  |::|  :: | : |::
VHFDWKKRLYVKFSLLLSVAYLQYTLTMLALTVALFFLLAYLPGIVPFFSVSLISYCHMRIVYAVLLKVEQHGGEPOKRS
      150      160      170      180      190      200      210

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 290

A DNA sequence (GBSx0318) was identified in *S.agalactiae* <SEQ ID 933> which encodes the amino acid sequence <SEQ ID 934>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1827(Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAC44392 GB:U43526 ORF-1 [Streptococcus pneumoniae]
    Identities = 48/151 (31%), Positives = 66/151 (42%), Gaps = 5/151 (3%)

    Query: 1  MIYDHLNLNTHYKDINPNLDLAIDYLLSHDLRNLDIGTYHISPEVILMVQSNQLES-FD 59
              MI + L Y +NP+ ID+L L NL G+ I + L++
    Sbjct: 1  MIITKISRLGTYVGVNPHFATLIDFLEKTIGLENLTEGSIAIDGNRLFGNCFTYLDGQAG 60

25    Query: 60  HIFEYHKKYLDIHYVIEGHEVIKLGKGDKEV-EEY--LGDIGFIKCSEETSFDLRDNYI 116
              FE H+KYLDIH V+E E + + + V V +EY DI E LR
    Sbjct: 61  AFFETHQKYLDIHLVLENEEAMAVTSPENVSVTQYEDEEKDIELYTGKVEQLVHLRAGEC 120

30    Query: 117 AFFFPFEEAHQPNMGSLGNVYKKGVLKVLMA 147
              FPE+ HQP + VKK V KV ++
    Sbjct: 121 LITFPEDLHQP-KVRINDEPVKKVVKVAIS 150

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 291

A DNA sequence (GBSx0319) was identified in *S.agalactiae* <SEQ ID 935> which encodes the amino acid sequence <SEQ ID 936>. This protein is predicted to be sugar ABC transporter, permease protein (araQ).

- 40 Analysis of this protein sequence reveals the following:

```

    Possible site: 35
    >>> Seems to have a cleavable N-term signal seq.

    INTEGRAL Likelihood = -7.38 Transmembrane 245 - 261 ( 239 - 265)
    INTEGRAL Likelihood = -3.72 Transmembrane 140 - 156 ( 139 - 158)
45    INTEGRAL Likelihood = -3.61 Transmembrane 76 - 92 ( 71 - 94)
    INTEGRAL Likelihood = -2.81 Transmembrane 112 - 128 ( 107 - 128)
    INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 ( 186 - 204)

    ----- Final Results -----
50    bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-369-

>GP:AAD35515 GB:AE001721 sugar ABC transporter, permease protein

[Thermotoga maritima]

Identities = 94/262 (35%), Positives = 158/262 (59%), Gaps = 1/262 (0%)

5 Query: 15 LILCLLTVLFIFFYWIMTGAPKSPQDTIIIPPQWWPKAPTLENFKALTQVNPALRWLWN 74
 + + + V+F+ P ++ + +FK + PP +PK P+LE + + + L +L N
 Sbjct: 9 IFIVFMLVVFMLPVFVAVVSSFKPMSEIYSYPPTIFPKKPSLEGYINVIKEYDLLTYLRN 68

10 Query: 75 SVFISIMTMFLVCCTSSMAGYVLAKKRFYQKILFSLFIAAMALPKQVVLVPLVRIINFM 134
 ++F++ + + S M GY LAK +F+G + + S+F M + QV++VPL +I +
 Sbjct: 69 TLFVATVATVITVLVSMTGYGLAKGKFWGIRPVNSMFMFMVSAQVIMVPLFVVIRSL 128

15 Query: 135 GIHDTLWAVILPLVGWPFVFLMKQFSENIPTELLES AKIDGCGEIRTFINVAFFIVKPG 194
 G+ ++LW +I+P V P G+F+ Q+ ++IP ELLESAKIDG E + F + FF+ KP
 Sbjct: 129 GLINSLWGLIIPAVYTFTGMFMAVQYMKDIPDELLES AKIDGANEWQIFWRIVFPLSKPL 188

20 Query: 195 FAALAIFTFINTWNDYFMQLVMLTSRNNLTISLGVATMQAEM-ATNYGLIMAGAALAAVP 253
 AALAIF+F WND+ + L+++ RN T+ L +AT+Q E + I+A + L +P
 Sbjct: 189 VAALAIFGFTWRWDFVLPLLVRNRNLYTLQLALATIQEYGGAEWNTILAFSTLTIIIP 248

Query: 254 IVTVFLVFQKSFTQGITMGAVK 275
 + +FL+EQ+ F +GI G +K
 Sbjct: 249 TLIIFLLFQRLFMKGIMAGGLK 270

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 937> which encodes the amino acid sequence <SEQ ID 938>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -6.37 Transmembrane 245 - 261 (240 - 265)
 INTEGRAL Likelihood = -5.15 Transmembrane 140 - 156 (139 - 158)
 INTEGRAL Likelihood = -2.97 Transmembrane 111 - 127 (107 - 128)
 INTEGRAL Likelihood = -2.87 Transmembrane 76 - 92 (75 - 93)
 INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 (186 - 204)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 40 The protein has homology with the following sequences in the databases:

>GP:CAB59597 GB:AL132662 probable sugar transport inner membrane protein [Streptomyces coelicolor A3(2)]

Identities = 88/262 (33%), Positives = 147/262 (55%)

45 Query: 15 VMLCVLTILFIFFYWIMTGAFKAQADTIMIPPQWWPKAPTIENTFKALVVQNPALKWLWN 74
 ++L L ++F P W++ + + A+ PP WP + ++ ++ +W N
 Sbjct: 38 LLLAPLALVFAVPLVWLVLSSVMSNABINRFPFALWPSGIDLGGYRYVLGNAMFPRWFVN 97

50 Query: 75 SVFISVATMFLVCGTSSLAGYALAKKRFYQRLLESIFIAAMALPKQVVLVPLVRIINFM 134
 S+ +S T+ SLAGYA A+ RF G R+L + +A MA+P Q+ ++P ++ +
 Sbjct: 98 SLIVSAVTVAANLVFGSLAGYAFARMRFAGSRVLMGLMLATMAVPFQLTMIPTFLVMKKL 157

55 Query: 135 GIHDTLAAVILPLVGWPFVFLMKQFSENIPTELLES AKIDGCGEIRTFINVAFFIVKPG 194
 G+ DTL A+I+P + PF VFL++QF ++P EL E+A IDGC +R + + P+ +P
 Sbjct: 158 GLIDTLGALIVPSLVTPFAVFLLRQFFLSLPRELEAAWIDGCSRLRVLWRIVPLSRPA 217

60 Query: 195 FAALAIFTFINTWNDYFMQLVMLTSRENLTISLGVATMQAEMATNYGLIMAGAAMAAPVI 254
 A +A+ TF+ TWND L+ + T+ LG+ T Q + T + +MAG + +P+
 Sbjct: 218 LATVAVLTLFTTWNDLTWPLIAINHDTQYTLQLGLTTFQGHHTQWAAVMAGNVITVLPV 277

Query: 255 VTVELVFQKSFTQGITMGAVKG 276
 + FL QK+F Q IT +KG
 Sbjct: 278 LLAFLGAQKTFIQSITSSGLKG 299

- 65 An alignment of the GAS and GBS proteins is shown below:

-370-

Identities = 245/276 (88%), Positives = 262/276 (94%)

Query: 1 MKKKTFSAYNFLTALILCLLTVLFIFFPYWIMTGAFKSPDTIIPPQWWPKAPTLENFK 60
 M KK +A + LT ++LC+LT+LFIFPPYWIMTGAFK+Q DTI+IPPQWWPKAPT+ENFK
 5 Sbjct: 1 MTKKKLTASDILTTVMCLCVLTILFIFFPYWIMTGAFKAQADTIMIPPQWWPKAPTLENFK 60

Query: 61 ALTVQNPALRWLWNSVFISIMTFLVCCTSSMAGYVLAKKRFYQKILFSLFIAAMALPK 120
 AL VQNPAL+WLWNSVFIS+ TMFLVC TSS+AGY LAKKRFYQ++LFS+FIAAMALPK
 10 Sbjct: 61 ALVVQNPALKWLWNSVFISVATMFLVCGTSSLAGYALAKKRFYQRLFSIFIAAMALPK 120

Query: 121 QVVLVPLVRIINFMGIHDTLWAVILPLVGWPFVFLMKQFSENIPTELLESKIDGCGEI 180
 QVVLVPLVRI+NFMGIHDTL AVILPLVGWPFVFLMKQFSENIPTELLESKIDGCGEI
 Sbjct: 121 QVVLVPLVRIINFMGIHDTLAAVILPLVGWPFVFLMKQFSENIPTELLESKIDGCGEI 180

Query: 181 RTFINVAFPIVKPGFAALAIPTFINTWNYDFMQLVMLTSRNNLTISLGVATMQAEMATNY 240
 RTF NVAFPIVKPGFAALAIPTFINTWNYDFMQLVMLTSR NLTISLGVATMQAEMATNY
 15 Sbjct: 181 RTFFNVAFPIVKPGFAALAIPTFINTWNYDFMQLVMLTSRENLTISLGVATMQAEMATNY 240

Query: 241 GLIMAGAALAAPVIVTVFLVFQKSFTQGITMGAVKG 276
 GLIMAGAA+AAVPIVTVFLVFQKSFTQGITMGAVKG
 20 Sbjct: 241 GLIMAGAMAAPVIVTVFLVFQKSFTQGITMGAVKG 276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 292

A DNA sequence (GBSx0320) was identified in *S. agalactiae* <SEQ ID 939> which encodes the amino acid sequence <SEQ ID 940>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.

30	INTEGRAL	Likelihood = -10.83	Transmembrane	74 - 90 (64 - 96)
	INTEGRAL	Likelihood = -6.37	Transmembrane	108 - 124 (107 - 126)
	INTEGRAL	Likelihood = -5.84	Transmembrane	270 - 286 (265 - 290)
	INTEGRAL	Likelihood = -5.20	Transmembrane	161 - 177 (156 - 182)
35	INTEGRAL	Likelihood = -0.16	Transmembrane	219 - 235 (219 - 235)

----- Final Results -----

	bacterial membrane	---	Certainty=0.5331(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
40	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:EAB05584 GB:AP001513 sugar transport system (permease) (binding protein dependent transporter) [Bacillus halodurans]
 Identities = 106/289 (36%), Positives = 168/289 (57%), Gaps = 6/289 (2%)

45 Query: 9 RETMIAYAF LAPILLFFLI FVFAPVMVG FVTSFFNYSM-TQFTFIGLANYNRMF-HDSIF 66
 +E Y F+AP ++ F IF PM+ SF ++ + + + G NY R+F D +F
 Sbjct: 25 KEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLYNEPVWTGADNYYRLFVTDLLF 84

50 Query: 67 MKSLINTVIIVIGSVPVVVFSLFVAANTYKKNVFSRSFYRCVFFLPVVTGSAVTVVWK 126
 K++ NT +G +P+ + SL +A +K V + +R FFLP V+ VA+T++W+
 Sbjct: 85 RKTVFNTFYAALG-IPIGMAVSLGIAVALNOK-VKGIALFRTAFFLPAVSSVVAITLLWR 142

55 Query: 127 WIYDPMGSIILNYILKSGHVIEQNISWLGDKHWALLAIIIIILLTTSVGQPIILYIAAMGNI 186
 WI++ G+LN +L +V WL D+ WA+ A+II + +G +ILY+AA+ +
 Sbjct: 143 WIFNADFGLNIMLN--YVGIHGPGWLSDEKWAMPAMIIQGVWGLGINMILYLALQGV 200

Query: 187 DNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFAIQLLTSGGPNYS 246
 + +L EAA +DG N Q F I PS+ PTT +I + +TI + Q F ++T GGPNNYS
 60 Sbjct: 201 NPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITSTIGALQDFQRFMIMTEGGPNYS 260

Query: 247 TSTLMYYLYEKAPKLSEYGYANTMGVFLAVMIALISFAQPKILGNDVEY 295

-371-

T+T++YYL+ AF+ E GYA+ M L ++I +I+ FK+ V Y
 Sbjct: 261 TITVVYYLFLNAFRYMEMGYASAMAWVLGIILLIITIINFKLAKKVVHY 309

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 941> which encodes the amino acid sequence <SEQ ID 942>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -12.74 Transmembrane 55 - 71 (44 - 78)
 INTEGRAL Likelihood = -10.83 Transmembrane 109 - 125 (98 - 130)
 INTEGRAL Likelihood = -6.21 Transmembrane 304 - 320 (299 - 324)
 INTEGRAL Likelihood = -6.00 Transmembrane 142 - 158 (141 - 160)
 INTEGRAL Likelihood = -5.04 Transmembrane 196 - 212 (190 - 216)
 INTEGRAL Likelihood = -0.16 Transmembrane 253 - 269 (253 - 269)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the databases:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding
 protein dependent transporter) [Bacillus halodurans]
 Identities = 113/310 (36%), Positives = 176/310 (56%), Gaps = 9/310 (2%)
 25 Query: 25 KVEQKKEVFQVNVNKLKMR---ETLISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSM- 80
 +VE +E K K R E Y F+AP ++ F IF L PM+ SF ++ +
 Sbjct: 4 EVETPRETKTTKARKQKRLNKEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLY 63
 30 Query: 81 TEFTFVGFANYARMF-QDPIFMKSLINTLIIVIGSVPVVVFSLFVAAKTYDKNVVARSF 139
 E + G NY R+F D +F K++ NT +G +P+ + SL +A K V +
 Sbjct: 64 NEPVWTGADNYRLEFVTDDLFRKTVFNTFYAALG-IPIGMAVSLGIAVALNQK-VKGIAL 121
 35 Query: 140 YRAVFFLPVVTGSAVAVTVVWKWIYDPMGILNYVLKYAHVIEQNISWLGDKHWALLAIIV 199
 +R FFLP V+ VA+T++W+WI++ G+LN +L Y + WL D+ WA+ A+I+
 Sbjct: 122 FRTAFFLEPAVSSVVAITLLWRWIFNADFGLLNIMLNIVGI--HGPGWLSDEKWAMPAMII 179
 40 Query: 200 ILLTTSVGQPIILYIAAMGNIDNSLVEAARVDGATEFQVFWNIKWPSLLPTTLYIAVITT 259
 + +G +ILY+AA+ ++ +L EAA +DG +Q F +I PS+ PTF +I + +T
 Sbjct: 180 QGVWGGGLGINMILYLAALQGVNPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITST 239
 45 Query: 260 INSQCFCALIQLLTSGGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQ 319
 I + Q F ++T GGPNYST+T++YYL+ AF+ E GYA+ M L ++I II+
 Sbjct: 240 IGALQDFQRFMIMTEGGPNYSTTITVVYYLFLNAFRYMEMGYASAMAWVLGIILLIITIIN 299
 50 Query: 320 FKILGNDVEY 329
 FK+ V Y
 Sbjct: 300 FKLAKKVVHY 309

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 263/295 (89%), Positives = 278/295 (94%)
 Query: 1 MRTNKLKMRETMIAYAFAPILLFFLIIFVAPMVMGFVTSFFNYSMTOFTFIGLANYNRM 60
 + NKLKMRET+I+YAFAP+L+FF+IFV PM+MGFVTSFFNYSM+FTF+G ANY RM
 Sbjct: 35 VNVNKLKMRETLISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSMTEFTFVGFANYARM 94
 55 Query: 61 FHDSIFMKSLINTVIIVIGSVPVVVFSLFVAANTYKKNVFSRSFYRCVFFLPVVTGSVA 120
 F D IFMKSLINT+IIVIGSVPVVVFSLFVAA TY+KNV +RSFYR VFFLPVVTGSVA
 Sbjct: 95 FQDPIFMKSLINTLIIVIGSVPVVVFSLFVAAKTYDKNVVARSFYRAVFFLPVVTGSVA 154
 60 Query: 121 VTVVWKWIYDPMGILNYILKSGHVIEQNISWLGDKHWALLAIILLTTSVGQPIILYI 180
 VTVVWKWIYDPMGILNY+LK HVIEQNISWLGDKHWALLAI+ILLTTSVGQPIILYI
 Sbjct: 155 VTVVWKWIYDPMGILNYVLKYAHVIEQNISWLGDKHWALLAIIVILLTTSVGQPIILYI 214
 Query: 181 AAMGNIDNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSQCFCALIQLLTS 240

-372-

AAMGNIDNSL EAARVDGA E QVFW IKWPSLLPTTLYIAVITTTINSFQCFALIQLLTS
 Sbjct: 215 AAMGNIDNSLVEAARVDGATEFQVFWNIKWPSLLPTTLYIAVITTTINSFQCFALIQLLTS 274

Query: 241 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295
 5 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIA+ISFAQFKILGNDVEY
 Sbjct: 275 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQFKILGNDVEY 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 293

A DNA sequence (GBSx0321) was identified in *S. agalactiae* <SEQ ID 943> which encodes the amino acid sequence <SEQ ID 944>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 54/187 (28%), Positives = 90/187 (47%), Gaps = 14/187 (7%)

Query: 19 MFACVDSSQSVMAAEKD-KVEITWNAFPTFTQEKAKDGVGTYEKKVIAFEKKNPNKVK 77
 MF+ + + ++D + I WW + D Y KVI+ +EKNP++ ++
 Sbjct: 1 MFGSGSAGEEASGKKEDVTLRIAWG-----GQPRHD----YTTKVIELYEKKNPHVHIE 51

Query: 78 LETIDFTSGPEKITTAIEAGTAPDVLFDAPGRITQYKNGKGLADLNDLFTDQFIKDVN-- 135
 E ++ +K+ AG PDV+ + QY GK +L DL D I DV+
 Sbjct: 52 AEFANWDDYWKKLAPMSAAGQLPDVVIQMDTAYLAQYKGNQLEDLTPYTKDGTI-DVSSI 110

Query: 136 NKNIIQASKSGDKAYMYPISAPFYMAFNKKMLKDGVLKLVEGWTTSDFEKVLKALKN 195
 ++N++ K +K Y + + + N+ +LK AGV + +E WT D+EK+ L+
 Sbjct: 111 DENMLSGGKIDNKLYGFTLGVNVLVSIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQE 169

Query: 196 KGYTPGS 202
 K GS
 Sbjct: 170 KAGVYGS 176

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 945> which encodes the amino acid sequence <SEQ ID 946>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

!GB:Z99107 similar to sugar-binding protein [Bacillu... 82 2e-14

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 105/446 (23%), Positives = 176/446 (38%), Gaps = 71/446 (15%)

Query: 24 GKSQKEAGASKSDTAKTEITWNAFVFTQEKAEKDGVTYKKLIAAFEKANPEIKVKLET 83
 G S E + K + I WW + D Y K+I +EK NP + + + E

-373-

Sbjct: 4 GCSAGEASGKKEDVTLRIAWWG-----GQPRHD----YTTKVIELYEKKNPHVHIEAEF 54

Query: 84 IDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNDLFTTEFTKDVN--NDK 141
 ++ +K+ AG PDV+ + QY GK +L DL +T++ T DV+ ++

5 Sbjct: 55 ANWDDYWKKLAPMSAAGQLPDVIMDTAYLAQYGGKNQLEDLTP-YTKDGTIDVSSIDEN 113

Query: 142 LIQASKAGDTAYMYPISAPFYMALNKKMLKDAGVLDLVKEGWTDDFEKVLKALKDK-- 199
 ++ K + Y + + + N+ +LK AGV + +E WT +D+EK+ L++K

10 Sbjct: 114 MLSSGKIDNKLYGFTLGVNVLSEIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQEKAG 172

Query: 200 -----GYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKV-----TKYTT 239
 G +P F +G R + + DD++ T T

Sbjct: 173 VYGSNGMHPPDIFPPYLRKTGERFVKEDGTGLAYQDDQLFVDYFERQLRLVKAKTSPTP 232

15 Query: 240 DDANSIKAMTKISNWKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLE 299
 D++ IK M +D ++ G SA N++N F A+L +

Sbjct: 233 DESAQIKGM-----EDDFIVKGGK----SAITWYNSNQYLGK-----ARLTD 269

Query: 300 ASKVDYLEIPFSPDDGKPELEYLVNGFAVFNKDEQKVAASKTFIQFIADDDKEWGPKNVV 359
 + YL P + L + E K A+K FI F +++E + +

20 Sbjct: 270 SPLSLYLP---PEQMGEKALTLKPSMLFSIPKSSEHKKEAAK-FINFFVNNEE-ANQLIK 324

Query: 360 RTGAFPVRTSYGDLYKDKRMEK---IAEWTKFYSPYNTID----GFAEMRTLWFPMVQ 411
 PV D K K E+ I E+ + S + D G AE+ L

25 Sbjct: 325 GERGVPSVDKVAIDAKPKLNEEETNIVEYVETASKNISKADPPPEVGSAEVIKLLKDTSD 384

Query: 412 AVSNGDEKPEDALKAFTEKANKTIKK 437
 + PE A K F +KAN+ +++

30 Sbjct: 385 QILYQKVSPEKAAKTFRKANEILER 410

An alignment of the GAS and GBS proteins is shown below:

Identities = 352/438 (80%), Positives = 384/438 (87%), Gaps = 4/438 (0%)

Query: 1 MSIKKSVIGFCLGAAALSMFACVDSSQSVMAAEKD---KVEITWWAFPTFTQEKADGVG 57
 M+++KK LGA+ L + AC SQ A K K EITWWAFP FTQEKADGVG

35 Sbjct: 1 MNMKKLASLAMLGASVLGLAACGGKSQKEAGASKSDTAKTEITWWAFPVFTQEKADGVG 60

Query: 58 TYEKKVIAFEKKNPNIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 117
 TYEKK+I AFEK NP IKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG

40 Sbjct: 61 TYEKKLIAAFEKANPEIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 120

Query: 118 KLADLNDLFTDQFIKDVNNKNIQASKSGDKAYMYPISAPFYMAFNKKMLKDAGVLKLV 177
 KLADLNDLFT++F KDVNN +IQASK+GD AYMYPISSAPFYMA NKKMLKDAGVL LV

45 Sbjct: 121 KLADLNDLFTTEFTKDVNNDKLIQASKAGDTAYMYPISAPFYMALNKKMLKDAGVLDLV 180

Query: 178 KEGWTTSDFEKVLKALKNGYTPGSFFANGQGGDQGPRAFFANLYSAPITDKEVTKYTTD 237
 KEGWTT DFEKVLKALK+KGY PGSFFANGQGGDQGPRAFFANLYS+ ITD +VTKYTTD

Sbjct: 181 KEGWTTDDFEKVLKALKDKGYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKVTKYTTD 240

50 Query: 238 TKNSVSKSMKKIVIEWIKKGYLMNGSQYDGSADIQNFANGQTAFTILWAPAQPKTQAKLLES 297
 NS+K+M KI WIK G +MNGSQYDGSADIQNFANGQT+FTILWAPAQ OAKLLE+

Sbjct: 241 DANSIKAMTKISNWKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLEA 300

Query: 298 SKVDYLEVPPFSEDGKPDLEYLVNGFAVFNKDKENKVKASKKFITFIADDDKKGPKDVIR 357
 SKVDYLE+PFPS+DGKP+LEYLVNGFAVFNKDE KV ASK FI FIADDDK+WGPK+V+R

55 Sbjct: 301 SKVDYLEIPFSPDDGKPELEYLVNGFAVFNKDEQKVAASKTFIQFIADDDKEWGPKNVVR 360

Query: 358 TGAFPPVRTSYGDLYKDKRMMKISKWTQYYSPPYNTIDGFSEMRTLWFPMVQSVSNGDEK 417
 TGAFPPVRTS+GDLYK DKRM KI++WT++YSPYNTIDGF+EMRTLWFPMVQ+VSNGDEK

60 Sbjct: 361 TGAFPPVRTSYGDLYK-DKRMEKIAEWTKFYSPYNTIDGFAEMRTLWFPMVQAVSNGDEK 419

Query: 418 PADALKDFTQKANDTIKK 435
 P DALK FT+KAN TIKK

65 Sbjct: 420 PEDALKAFTEKANKTIKK 437

[illegible]

-375-

SEQ ID 944 (GBS16) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 9; MW 49kDa).

The GBS16-His fusion product was purified (Figure 92A; see also Figure 189, lane 9) and used to immunise mice (lane 1 + 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 92B), FACS (Figure 92C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 294

- 10 A DNA sequence (GBSx0322) was identified in *S.agalactiae* <SEQ ID 947> which encodes the amino acid sequence <SEQ ID 948>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9459> which encodes amino acid sequence <SEQ ID 9460> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:AAC66999 GB:AE001166 conserved hypothetical protein [Borrelia burgdorferi]
25 Identities = 107/225 (47%), Positives = 147/225 (64%), Gaps = 6/225 (2%)
Query: 12 QIKNGIIVSCQALPGEPLYTESGGVMPILLALAAQFAGAVGIRANSVRDIKEIQEVTNLPI 71
+IK G+IVSCQAL EPL+ S +M +ALAA+ GA+GIRAN V DI +I+ +LPI
Sbjct: 6 KIKRGLIVSCQALENEPLH--SSFIMSKMALAAKIGGAIGIRANGVNDISQIKLEVDLPI 63
30 Query: 72 IGI IKREYPPQEPFITATMTTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQIKRKYP 131
IGI IK+ Y + FIT TM E+D+L + + +IALD T R R DG+ + +F + IK+KYP
Sbjct: 64 IGI IKKNYNKCDVFITPTMKEDLCNEGVDITIALDATFRNRPDGVLLDDFFENIKKKYP 123
35 Query: 132 EQLLMADISTFEEGKNAFEAGVDFVGTTLISGYTDYSR--QEEGPDIELLNKLCQAGI--D 187
+Q LMADIS+ +E NA + G DF+GTTL GYT + D L L + +
Sbjct: 124 KQCLMADISSLDEAINADKLGFDFTTLYGYTKNTNGLNIADNDFNFLTLLNSNLKST 183
40 Query: 188 VIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGLS 232
+I EGKI TP +A + +GV +VVGGAITRP EI ++F+ ++
Sbjct: 184 LIVEGKIDTPLKAQKCFEMGVDLVVVGGAITRPAEITKKFVEKIN 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 949> which encodes the amino acid sequence <SEQ ID 950>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.49 Transmembrane 175 - 191 (175 - 192)
50 ----- Final Results -----
bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-376-

The protein has homology with the following sequences in the databases:

```

>GP:AAD28762 GB:AF130859 putative N-acetylmannosamine-6-P epimerase
[Clostridium perfringens]
Identities = 113/225 (50%), Positives = 148/225 (65%), Gaps = 5/225 (2%)

5   Query: 10 LMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQAITD 69
      +++ +KG +IVSCQAL EPL+S IM MA AA++ GA IRA + DI EI+ +T
      Sbjct: 1 MLDVVKGNLIVSCQALSDEPLHSSF--IMGRMAIAAKQGGAAAIRAQGIDDINEIKEVTK 58

10  Query: 70 LPIIGIHKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKDRHDGLDIASFIRQVKE 129
      LPIIGIHK++Y E +IT TM EVD+L + +I +D TKR R +G +I + +
      Sbjct: 59 LPIIGIHKRYDDSEIYITPTMKEVDELLKTDCEMIGLDATKRKRPNGENIKDLVDIHA 118

15  Query: 130 KYPNQLLMADISTFDEGLVAHQAGIDFVGTTLGGYTPYSRQEAGPDVALIEALCK-AGIA 188
      K +L MADIST +EG+ A + G D V TTLGGYTPYS+Q D L+E L K I
      Sbjct: 119 K--GLAMADISTLEEGIEAKLGFDCVSTTLGGYTPYSKQNSVDFELLEELVKTKVIP 176

      Query: 189 VIAEGKIHSPEEAKKINDLGAVGVGGAITRPKEIAERFIEALK 233
      VI EG+I++PEE KK DLG VVGGAITRP++I +RF + LK
20  Sbjct: 177 VICEGRINTPEELKKALDLGAYSAVVGGAITRPQQITKRFIDILK 221

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 172/227 (75%), Positives = 202/227 (88%)

25  Query: 5 SKEAFKKQIKNGIIVSCQALPGEPLYTESGGVMPLALAAQEAGAVGIRANSVRDIKEIQ 64
      +KE +Q+K GIIIVSCQALPGEPLY+E+GG+MPL+A AAQEAGAVGIRANSVRDIKEIQ
      Sbjct: 6 TKEKLMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQ 65

30  Query: 65 EVTNLPIIGIHKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQ 124
      +T+LPIIGIHK++YPPQEPFITATMTEVDQLA+L+IAVIA+DCT R+RHDGL + FI+
      Sbjct: 66 AITDLPIIGIHKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKDRHDGLDIASFIR 125

      Query: 125 KIKRKYPEQLLMADISTFEEGKNAFEAGVDFVGTTLGGYTDYSRQEAGPDIELLNKLCQA 184
      ++K KYP QLLMADISTF+EG A +AG+DFVGTTLGGYT YSRQE GPD+ L+ LC+A
35  Sbjct: 126 QVKEKYPNQLLMADISTFDEGLVAHQAGIDFVGTTLGGYTPYSRQEAGPDVALIEALCKA 185

      Query: 185 GIDVIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGL 231
      GI VIAEGKIHP++A +IN +GVAGIVVGGAITRPKEIAERFI L
40  Sbjct: 186 GIAVIAEGKIHSPEEAKKINDLGAVGVGGAITRPKEIAERFIEAL 232

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 295

A DNA sequence (GBSx0323) was identified in *S.agalactiae* <SEQ ID 951> which encodes the amino acid sequence <SEQ ID 952>. This protein is predicted to be group B streptococcal surface immunogenic protein. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

50  ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 953> which encodes the amino acid sequence <SEQ ID 954>. Analysis of this protein sequence reveals the following:

Possible site: 25

-377-

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 182/437 (41%), Positives = 240/437 (54%), Gaps = 53/437 (12%)
 Query: 1 MKMNKKVLLTSTMAASLLSVASVQAQETDITWTARTVSEVKADLVKQDNKSSYTVKYGDT 60
 M + KK L +++A SL+ +A+ QAQE WT R+V+E+K++LV DN +YTVKYGDT
 Sbjct: 1 MIITKKSFLVTSVALSLVPLATAQAQE----WTPRSVTEIKSELVLVDNVFTYTVKYGDT 56
 15 Query: 61 LSVISEAMSIDMNVLAKINNIADINLIYPETTLTVTYDQKSHITATSMKIETPATNAAGQT 120
 LS I+EAM ID++VL IN+IA+I+LI+P+T L' Y+Q AT++ ++ PA++ A +
 Sbjct: 57 LSTIAEAMGIDVHVLGDINHIANIDLIFPDILTANYNQHGQ-ATNLTVQAPASSPASVS 115
 20 Query: 121 TATVDLKTINQVSVADQKVSINTISEGMTP-EAATTIVSPMKTYSSAPALKSKEVLAQEQA 179
 Q S Q ++ TP + TT + K SS A S E+ +
 Sbjct: 116 HVPSSSEPLPQASATSQPTV--PMAPPATPSDVPTTPFASAKPDSVTA--SSELTSSSTND 171
 25 Query: 180 VSQAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAP 239
 VS ++E V P A E T V T +S A +A P P +
 Sbjct: 172 VSTELSSSQKQPEVPQEA VPTPKAAE-----TTEVEPKTDISEAPTSANRPVPNESASE 226
 30 Query: 240 VRTVAAPRVASVKVVTPKVETGASPEHVSAPAVP---VTTTSPATDSKLOATEVKSVPVA 296
 + AAP + A E SAPA TTS AT + L
 Sbjct: 227 EVSSAAP-----AQAPAEKEETSAPAAQKAVADTTSVATSNGL----- 264
 35 Query: 297 QKAPTATPVAQPASTTNVAHAHPENAGLQPHVAAYKEKVASTYGVNEFSTYRAGDPGDHG 356
 AP A +P NAGLQF AA+KE+VAS +G+ FS YR GDPGDHG
 Sbjct: 265 SYAPNH-----AYNPMNAGLQPTAAFKEEVASAFGITSFSGYRFGDPGDHG 311
 40 Query: 357 KGLAVDFIVGTNQALGNKVAQYSTQNMAANNISYVIWQOKFYSNINSYGPANTWNAMPD 416
 KGLA+DF+V N ALG++VAQY+ +MA ISYVIW+Q+FY+ SIYGPA TWN MPD
 Sbjct: 312 KGLAIDFMPENSALGDQVAQYAIIDHMAERGISYVIWKORFYAPFASIYGPAYTWNMPMD 371
 45 Query: 417 RGGVTANHYDHHVHVSFN 433
 RG +T NHYDHHVHVSFN
 Sbjct: 372 RGSITENHYDHHVHVSFN 388

A related GBS gene <SEQ ID 8539> and protein <SEQ ID 8540> were also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 3
 SRCFLG: 0
 McG: Length of UR: 20
 Peak Value of UR: 1.96
 Net Charge of CR: 2
 50 McG: Discrim Score: 2.95
 GvH: Signal Score (-7.5): 3.84
 Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 24
 55 ALOM program count: 0 value: 4.29 threshold: 0.0
 PERIPHERAL Likelihood = 4.29 58
 modified ALOM score: -1.36
 *** Reasoning Step: 3
 60 Rule gpol

----- Final Results -----

65 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-378-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8540 (GBS322) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 9; MW 52kDa). The GBS322-His fusion product was purified (Figure 214, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 267), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 296

A DNA sequence (GBSx0324) was identified in *S.agalactiae* <SEQ ID 955> which encodes the amino acid sequence <SEQ ID 956>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 (4 - 21)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC46072 GB:U50357 zoocin A endopeptidase [Streptococcus
 zooepidemicus]
 Identities = 163/274 (59%), Positives = 196/274 (71%), Gaps = 11/274 (4%)
 Query: 25 VLADTYVRPIDNGRITTTGFNGYPGHGVDYAVPTGTIIIRAVADGTVKFAGAGANFSWMTD 84
 V A TY RP+D G ITTGFNGYPGH GVDYAVP GT +RAVA+GTVKFAG GAN WM
 Sbjct: 21 VSAATYTRPLDTGNITTTGFNGYPGHVGVVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLW 80
 Query: 85 LAGNCVMIQHADGMHSGYAHMSRVVARTGEKVKQGDIIIGYVGATGMATGPHLHFEFLPAN 144
 +AGNCV+IQHADGMH+GYAH+S++ T VKQG IIGY GATG TGPPLHFE LPAN
 Sbjct: 81 MAGNCVLIQHADGMHTGYAHLKISVSTDSTVKQQLIIGYTGTATGQVTGPHLHFEMLPAN 140
 Query: 145 PNFQNGFPHGRINPTSLIANVATFSGKTQASAPSIKPLQSAVQVQSSKLVYRVDELQKV 204
 PN+QNGF GRI+PT IAN F+G T + P N LK+Y+VD+LQK+
 Sbjct: 141 PNFQNGFSGRIDPTGYIANAPVFNGTTPTE-----PTTPTTN----LKIYKVDLQKI 189
 Query: 205 NGVVLVKNNTLTPTGFDWNDNGIPASEIDEVDANGNLTAQVLQKGGYFI FNPKTLKTV 264
 NG+W V+NN L PT F W DNGI A ++ EV +NG T+DQVLQKGGYF+ NP +K+V
 Sbjct: 190 NGIWQVRNINLVPTDFTWVDNGIAADVIEVTSNGTRTSDQVLQKGGYFVINPNNVKS 249
 Query: 265 KPIQGTAGLTWAKTRFANGSSVWLRVDNSQELLY 298
 P++G+ GL+WA+ F G +VWL + LLY
 Sbjct: 250 TPMKSGGLSWAQVNFTTGGNVWLNTTSKDNLLY 283

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8541> and protein <SEQ ID 8542> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 6.63
 GvH: Signal Score (-7.5): -2.97
 Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -1.86 threshold: 0.0
 INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 (4 - 21)
 PERIPHERAL Likelihood = 5.57 50

-379-

modified ALOM score: 0.87

*** Reasoning Step: 3

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1744(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

GP|2804351|gb|AAC46072.1||U50357(21 - 283 of 285) zoocin A endopeptidase {Streptococcus
zooepidemicus}

%Match = 34.2

%Identity = 61.3 %Similarity = 74.4

15 Matches = 163 Mismatches = 65 Conservative Sub.s = 35

144 174 204 234 264 294 324 354
VV*VFLS*LRYTTYILKTFLEFKPPKYSSR*VLFLIF*FKFSNKLIASV*ALHYINSIWRFFLNKWLKASSLVVLGGMV20 MKRIFFAFPLSLCLF
1025 384 414 444 474 504 534 564 594
LSAGSRVLADIYVRPIDNGRITITGFNGYPGHGCGVDYAVPTGTIIIRAVADGTVKFAGAGANFSWMTDLAGNCVMIQHADGM
:
IFGTQTVSAATYTRPLDTGNITITGFNGYPGHVGVVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLWMAGNCVLIQHADGM
30 40 50 60 70 80 9030 624 654 684 714 744 774 804 834
HSGYAHMSRVVARTGEKVKQGDIIIGYVGATGMATGPHLHFEFLPANPNFQNGFHFGRINPTSLIANVATFSGKTQASAPSI
:
HTGYAHLISKISVSTDSTVKQGGIIGYTGATGQVTGPHLHFEMLPANPNWQNGFSGRIDPTGYIANAPVFNGTT-----
110 120 130 140 150 16035 864 894 924 954 984 1014 1044 1074
KPLQSAFVQNSKLVYRVDLQKVGWLVKNNTLTPTGFDWNDNGIPASEIDEVDANGNLTADQVLQKGGYFIENPK
:
-P--TEP-TTPTTNLKIYKVDLQKINGIWQVRNNILVPTDFTWVDNGIAADDVIEVTSNGTRTSDQVLQKGGYFVINPN
180 190 200 210 220 230 24040 1104 1134 1164 1194 1224 1254 1284 1314
TLKTVEKPIQGTAGLTWAKTRFANGSSVWLVRVDNSQELLYK*FEVLIHCFK*QLCY*LSTISLNLKIL*SSKV*YYSL
:
NVKSVGTPMKSGGLSWAQVNFTTGGNVWLNTTSKDNLLYGK
45 260 270 280SEQ ID 8542 (GBS36) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 4; MW 34.1kDa).

GBS36-His was purified as shown in Figure 192, lane 7.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 297A DNA sequence (GBSx0325) was identified in *S.agalactiae* <SEQ ID 957> which encodes the amino acid sequence <SEQ ID 958>. This protein is predicted to be phosphoribosylaminoimidazolecarboxamide
55 formyltransferase/IMP cyclohyd. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

-380-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2815(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA04352 GB:AP001509 phosphoribosylaminoimidazolecarboxamide
 formyltransferase/IMP cyclohydrolase [Bacillus halodurans]
 Identities = 310/515 (60%), Positives = 390/515 (75%), Gaps = 4/515 (0%)

10

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDVDTGFPEMMD 60
 M +RAL+SVS+K GI+ FAK L +I+STGGTK AL +AG+ I DVTGFPE++D
 Sbjct: 1 MKRRALISVSNKEGIVPFAKALVEHEVEIVSTGGTKRALQEAGIPVTGISDVTGFPEILD 60

15

Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAAKDNNEIIDLVVVNLYPFKETILRPDVTYDLAV 120
 GRVKTLHPNIHGGLLA R+ D HL +++I ID VVVNLYPF++TI +P+ T+ A+
 Sbjct: 61 GRVKTLHPNIHGGLLAMRERDEHLAQLNEHHIRPIDFVVVNLYPFQQTIAKPEATFADAI 120

20

Query: 121 ENIDIGGFSMLRSAAKNHASVTVVVDSADYATVLGELADASQTTFKTRQRLAAKAFRHTA 180
 ENIDIGGFSMLR+AAKNH VVVVD DY TVL ELAD +T++RLAAK FRHTA
 Sbjct: 121 ENIDIGGFSMLRAAKNHQHVTVVVDPVDYETVLKELADQGNVATETKRRLAAKVFRHTA 180

25

Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240
 AYDA+IAEY T VGE PE LT+T++ KQ +RYGENP Q A FYQK L SIA AKQ
 Sbjct: 181 AYDAMIAEYLTDAVGEESPESLTVTFEKKQDLRYGENPHQKATFYQKPLGAKASIAHAKQ 240

30

Query: 241 LNKELSFNNIRDADAAIRIIRDFKDSPTVVALKHMNPOGIGQADDIETANDYAYEADPV 300
 L+GKELS+NNI DADAA+ I+++FK+ P VA+KHMNPG+G + I+ A+D AYEADPV
 Sbjct: 241 LHKELSYNNINDADAALSIVKEFKE-PAAVAVKHMNPGVGTGETIKEAFDKAYEADPV 299

35

Query: 301 SIFGGIVVLNREVDAATAEKMHPIFLEIIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
 SIFGGI+ LNREVD TA+ + IFLEIIIIAPS+SEAL +LT+ KKNLR+L LP + +
 Sbjct: 300 SIFGGIIALNREVDVETAKTLKEIFLEIIIIAPSFSEALDVLTS-KKNLRLLTLPLNEE- 357

40

Query: 361 ASEVEAEYTGUVGGLLVQNDVVAENPSDWQVVTDROPTEQEATALEFAWKAIKYVKSNG 420
 ++ E T+ GG LVQ +D ++ ++ T R+PTE E AL+ AW+ +K+VKS
 Sbjct: 358 -NQAERKRTSIHGGLVQREEDTYGFEEAEIKIPTKREPTAEWEALKLAWRVVVKHVKNSA 416

45

Query: 421 IIIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFFPADNIEEIAAAGIK 480
 I++ + MT+G+GAGQ NRVG+ KIAIEQA + G+V+ SDAFFP D +E A AGI
 Sbjct: 417 IVLADGQMTVGVGAGQMNRVGAAKIAIEQAGEKAAGSVMSGDAFFPMGDTVELAAKAGIT 476

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMTFTGVRHFRH 515
 AIIQPGGS+RD+ESI+ A+KHG+ M+FTGVRHF+H
 Sbjct: 477 AIIQPGGSIRDEESTENADKHGAMVFTGVRHFH 511

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 959> which encodes the amino acid sequence <SEQ ID 960>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2932(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below:

Identities = 500/515 (97%), Positives = 507/515 (98%)

60

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDVDTGFPEMMD 60
 MTKRALISVSDKSGI+DFAKELKNLGWDIISTGGTKV LDDAGVETIAIDVDT GFPEMMD
 Sbjct: 1 MTKRALISVSDKSGI+DFAKELKNLGWDIISTGGTKVTLDDAGVETIAIDVTRFPEMMD 60
 Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAAKDNNEIIDLVVVNLYPFKETILRPDVTYDLAV 120

-381-

GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPD+TYDLAV
 Sbjct: 61 GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPDITYDLAV 120

Query: 121 ENIDIGGPSMLRSAAKNHSVTVVDSADYATVLGELADASQTTFKTRQLAAKAFRHTA 180
 ENIDIGGPSMLRSAAKNHSVTVVDV ADYATVLGELADA QTTF+TRQLAAK FRHTA
 Sbjct: 121 ENIDIGGPSMLRSAAKNHSVTVVDPADYATVLGELADAGQTTFETRQLAAKVFRHTA 180

Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQADDFYQKALPTDYSIASAKQ 240
 AYDALIAEYFT QVGEAKPEKLTITYDLKQAMRYGENPQQADDFYQKALPTDYSIASAKQ
 Sbjct: 181 AYDALIAEYFTTQVGEAKPEKLTITYDLKQAMRYGENPQQADDFYQKALPTDYSIASAKQ 240

Query: 241 LNGKELSFNNIRDADAAIRIIRDFKDSPTVVALKHMNPGGIGQADDIETAWDYAYEADPV 300
 LNGKELSFNNIRDADAAIRIIRDFKD PTVVALKHMNPGGIGQADDIETAWDY Y+ADPV
 Sbjct: 241 LNGKELSFNNIRDADAAIRIIRDFKDRPTVVALKHMNPGGIGQADDIETAWDYTYKADPV 300

Query: 301 SIFGGIVVLNREVDAAATAEKMHPIFLEIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
 SIFGGI+VLNREVDAAATA+KMHPIFLEIIAPSYSEEALAILTNKKKNLRILELPFDAQA
 Sbjct: 301 SIFGGIIVLNREVDAAATAKMHPIFLEIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360

Query: 361 ASEVEAEYTG VVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG 420
 ASEVEAEYTG VVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG
 Sbjct: 361 ASEVEAEYTG VVGGLLVQNQDVVAENPSDWQVVTDRQPTQEATALEFAWKAIKYVKSNG 420

Query: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK 480
 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK
 Sbjct: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK 480

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515
 AIIQPGGSVRDQ+SIDAANKHGLTMIFTGVRHFRH
 Sbjct: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 298

A DNA sequence (GBSx0326) was identified in *S. agalactiae* <SEQ ID 961> which encodes the amino acid sequence <SEQ ID 962>. This protein is predicted to be similar to antibiotic resistance protein. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1842 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12342 GB:Z99106 similar to antibiotic resistance protein
 [Bacillus subtilis]
 Identities = 65/263 (24%), Positives = 117/263 (43%), Gaps = 34/263 (12%)

Query: 5 KNLEIVESIFGD-WDETIWSCV-QGIMGEVFDLSLDQPKSSLAKLGRKSSFGFLAGQPT 62
 K ++++F D + T ++S + Q I G V+ D PKS +G +S F+AG
 Sbjct: 10 KKYSSLKTMFDDKYCPTFFVYSILDQTIPGAVYADDQTFPKSFF--IGTESGIYFIAGDQG 67

Query: 63 -----LFLLEVCSGEDIILVPQHKGWSDLIESTYQNAHSFKRYATKKDTLFERS 112
 + +V S + L W +++ + + +R A +
 Sbjct: 68 NRDFHDFIAGYEEQVKSSKRFTLFSSSDTWDSVLKPKLDDLNQMRRAAFSY----QP 122

Query: 113 RLEKFVTQLPNGFELRAIDEKV-----YNSCLEKEWSQDLVANYATYQYKKQGIGYVV 166
 + K QLP G L+ IDE + +NS +E+ + + + +G G+ V
 Sbjct: 123 KSFKKTLQLPKGLVLKRIDEIISHSTAFNSAYEEY-----WNSVSQFASKGFGFAV 175

-382-

Query: 167 YYQGNIIAGASSYSTYKNGIEIEVDTHPDFRRRLATIVAAQLILTCLDKGIYPSWDAH- 225
 + ++++ +S N E+++ T ++R GLA VA + I C++ GI PSWD
 Sbjct: 176 LHGNHVSECTSI FLGHNRAEMDIYTL EEYRGLGLAYCVANRFIAFCMENGIVPSWDCCI 235

Query: 226 -TRTSLNLSEKLGYEFSHEYIAY 247
 +S+ L+ KLG++ EY Y
 Sbjct: 236 CNNSSIALAAKLGFKTVTEYTIY 258

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 299

A DNA sequence (GBSx0328) was identified in *S.agalactiae* <SEQ ID 963> which encodes the amino acid sequence <SEQ ID 964>. This protein is predicted to be phosphoribosylglycinamide formyltransferase homolog (purN). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0736(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 965> which encodes the amino acid sequence <SEQ ID 966>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.53 Transmembrane 75 - 91 (75 - 91)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35

The protein has homology with the following sequences in the databases:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

40 Query: 50 LKLLVVGSGGREHAIAKLLASKGVDQVFPVAPCNDGMTLDGLDLVNIVVSEHSRLIAFAK 109
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+
 Sbjct: 1 MKILVIGSGGREHALAKFMESPOVEEVFPVAPGNSGMEKDGIGIIVHISELSNDKLVKFAQ 60

45 Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
 I F+GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNI GLTFVGFETALMNGVVDFAFIKAE LPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

50 Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
 Y TF E A AY++E+G P+V+KADGLA GKG V VA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGEESLFAFANGDKFYIMPTAQDHKRAFDGDKGPNITGGMGAYAPVPHLPQ 289
 +VVIEEFLDGEESLFAFANGDKFYIMPTAQDHKRAFDGDKGPNITGGMGAYAPVPHLPQ
 Sbjct: 176 GKVVIIEEFLDGEESLFAFANGDKFYIMPTAQDHKRAFDGDKGPNITGGMGAYSEVHLHISK 235

55

Query: 290 SVVDITAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNRFQDPETQIILPR 349
 VV+ A+E +V+P + GM+ EG+ + GVLY GLILT DG K IEFN+RFQDPETQ++LPR
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

-383-

Query: 350 LTSDFAQNIDIMMGIEFYITWQKDGVTLGVVVASEGYFPDYEGVPLPEKTDGDIITYY 409
 L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP + G+ LPE +G + YY
 Sbjet: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTLGVVVAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDITGLFYRNDIGSKAI 468
 AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI
 Sbjet: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 172/182 (94%), Positives = 176/182 (96%)

Query: 1 MKIAVFASGNGSNFQVIAEQFQVSFVFS DHRDAYVLER AQNLAIPSF AFELKEFENK AAY 60
 MKIAVFASGNGSNFQVIAEQF VSFVFS DHRDAYVLER AQNLAIPSF AFELKEFENK AY
 15 Sbjet: 1 MKIAVFASGNGSNFQVIAEQFPVSFVFS DHRDAYVLER AQNLAIPSF AFELKEFENK VAY 60

Query: 61 EQAVVDLLDKHEIDLVLCLAGYMKIVGETLLSAYEGRIINIHTYLP EFPGAHGIKDAWEA 120
 EQA+VDLLDKHEIDLVLCLAGYMKIVGETLL AYE RIINIHP YLPEFPGAHI+DAWEA
 20 Sbjet: 61 EQAIVDLLDKHEIDLVLCLAGYMKIVGETLLLAYERRIINIHPAYLPEFPGAHI+DAWEA 120

Query: 121 GVDQSGVTIHWVDSGVDTGQVIQQVHVPRLADDSLESFETRIHETEQLYPAVLDSLGIK 180
 GVDQSGVTIHWVDSGVDTGQVIQQV VPRLADDSLESFETRIHETEQLYPAVLDSLGI++
 Sbjet: 121 GVDQSGVTIHWVDSGVDTGQVIQQVRVPRLADDSLESFETRIHETEQLYPAVLDSLGE 180

25 Query: 181 RK 182
 RK
 Sbjet: 181 RK 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 30 vaccines or diagnostics.

Example 300

A DNA sequence (GBSx0329) was identified in *S.agalactiae* <SEQ ID 967> which encodes the amino acid
 sequence <SEQ ID 968>. Analysis of this protein sequence reveals the following:

Possible site: 52
 35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 121 - 137 (121 - 137)

----- Final Results -----
 40 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC16901 GB:AF016634 phosphoribosylformylglycinamide
 45 cyclo-ligase [Lactococcus lactis subsp. cremoris]
 Identities = 253/338 (74%), Positives = 288/338 (84%), Gaps = 4/338 (1%)

Query: 4 KNAYASGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVLISGT 63
 +NAYA+SGVDVEAGYEVV RIKKHVA+TER GV+GALGGFGG FDL S VKEPVLISGT
 50 Sbjet: 5 ENAYAKSGVDVEAGYEVVSRIKKHVAKTERLGLV GALGGFGGS FDL SVLDVKEPVLISGT 64

Query: 64 DGVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQVVA 123
 DGVGTKLMLAI+ DKHDTIG DCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLEQVVA
 55 Sbjet: 65 DGVGTKLMLAIRADKHDTIGIDCVAMCVNDIIAAGAEPLYFLDYIATGKNIPEKLEQVVA 124

Query: 124 GVAEGCVQASALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSK-VKEGDILLG 182
 GVAEGC+QA AALIGGETAEMPGMY EDDYDLAGFAVGVAEKSQ+IDG K V+ GD+LLG
 Sbjet: 125 GVAEGCLQAGAALIGGETAEMPGMYDEDDYDLAGFAVGVAEKSQIDGKDV EAGDVLLG 184

60 Query: 183 LASSGIHNSGYSLVRRVFADYTGDEVLP ELEGKQLKDVLLPEFTRIYVKAALPLIKEELVN 242
 LASSGIHNSGYSLVR+VFAD+ +E LPEL+ + L D LL PT+IYVK LPLIK+ +

-384-

Sbjct: 185 LASSGIHNSGYSLVRKVFADFDLNESELPELD-QSLIDTLLTPTKIYVKELLPLIKQNKIK 243

Query: 243 GIAHITGGGFIEENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMGVG 302
 GIAHITGGGF EN+PRMF + L+AEI E VLPFIKALEKYG IKHEEM+EIFNMG+G

5 Sbjct: 244 GIAHITGGGFHENLPRMFGNSLSAEIVEGSDVLPFIKALEKYGSIKHEEMYEIFNMGIG 303

Query: 303 LMLDVNPENVDVRVKELLDEPVYEIGRIKKADDSVVIK 340
 +++ V PEN +K+ L+ +EIG+++ + + VVIK

10 Sbjct: 304 MVIAPENAAALKKELN--AFEIGQMVNRQEAPVVIK 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 969> which encodes the amino acid sequence <SEQ ID 970>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3236(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

20

An alignment of the GAS and GBS proteins is shown below:

Identities = 321/340 (94%), Positives = 332/340 (97%)

25 Query: 1 MSEKNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVL 60
 MSEKNAYA+SGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLS+TGVKEPVL+

Sbjct: 1 MSEKNAYAKSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSKTGVKEPVLV 60

Query: 61 SGTGCVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQ 120
 SGTGCVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLE+

30 Sbjct: 61 SGTGCVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYIATGKNNPVKLEE 120

Query: 121 VVAGVAEGCVQASALIGGETAEMPGMYCEDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180
 VV+GVAEGCVQA AALIGGETAEMPGMYG+DDYDLAGFAVGVAEKSQIIDGSKVKEGDIL

35 Sbjct: 121 VVSGVAEGCVQAGALIGGETAEMPGMYGQDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180

Query: 181 LGLASSGIHNSGYSLVRRVFADYTGDEVLPLEEGKQLKDVLLLEPTRIYVKAALPLIKEEL 240
 LGLASSGIHNSGYSLVRRVFADYTG E+LPELEGKQLKDVLLLEPTRIYVKAALPLIKEEL

Sbjct: 181 LGLASSGIHNSGYSLVRRVFADYTGKELLPELEGKQLKDVLLLEPTRIYVKAALPLIKEEL 240

40 Query: 241 VNGIAHITGGGFIEENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300
 V GI HITGGGFIEEN+PRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG

Sbjct: 241 VKGIGHITGGGFIEENIPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300

45 Query: 301 VGLMLDVNPENVDVRVKELLDEPVYEIGRIKKADDSVVIK 340
 VGLML V+PENV+RVKELLDEPVYEIGRIKKAD SVVIK

Sbjct: 301 VGLMLAVSPENVNRVKELLDEPVYEIGRIKKADASVVIK 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 301

A DNA sequence (GBSx0330) was identified in *S.agalactiae* <SEQ ID 971> which encodes the amino acid sequence <SEQ ID 972>. This protein is predicted to be phosphoribosylpyrophosphate amidotransferase (purF). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1112(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>

-385-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP: AAD12627 GB:U64311 phosphoribosylpyrophosphate amidotransferase
[Lactococcus lactis]
Identities = 340/470 (72%), Positives = 404/470 (85%), Gaps = 6/470 (1%)

Query: 3 YEVKSLNEECGVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGLS 62
+E K+LNEECG+FG+WG+P AA++TYFGLH+LQHRGQEGAGI+ N+NGKL +R +GL++

10 Sbjct: 37 FEAKTLNEECGLFGVWGHDPDAARLT YFGLHALQHRGQEGAGILVNNNGKLNHRHGLGLVT 96

Query: 63 EVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFYKFHDGQFALCHNGNLTNAISS 122
EVF+++ +L+ LTG++AIGHVRYATAGSA+I NIQPF ++FHDG L HNGNLTN S

15 Sbjct: 97 EVFRHEKDLEELTGSSAIGHVRYATAGSANINNIQPFQFEFHDGSLGLAHNGNLTN AQS L 156

Query: 123 RKELEKQGAIFNASSDTEILMHLIRSHNPSFMGKVKKEALSTVKGGFAYLLMTEDKLI A 182
R ELEK GAIF+++SDTEILMHLIRSH+P FMG+VKEAL+TVKGGFAYL+MTE+ ++AA

Sbjct: 157 RCELEKSGAIFSSNSDTEILMHLIRSHHPEFMGRVKEALNTVKGGFAYLIMTENSIV A A 216

20 Query: 183 LDPNFAFRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYTDE 242
LDPN FRPLSIG+M NGA V++SETCAF+VVG A W++DV+PGE+I I+D GI D++TD

Sbjct: 217 LDPNGFRPLSIGKMSNGALVVASETCAFVVGATWIDVQVGEIIEINDDGIHV DQFTDS 276

Query: 243 TQLAICSM EYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAAMG 302
T + ICSMEY+YFARPDS I GNVHTARK GK LAQE K DADIVIGVPNSSLSA A G

25 Sbjct: 277 TNMTICSM EYIYFARPDSNIAGVNVHTARKRSGKILAQEAKIDADIVIGVPNSSLSA ASG 336

Query: 303 FAESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGRVVMIDDSIVR 362
+AEESGLP EMGL+KNQY RTFIQPTQELREQGVRMKLSAV GVV+GKRV+M+DDSI VR

30 Sbjct: 337 YAEESGLPYEMGLIKNQYVARTFIQPTQELREQGVRMKLSAVRGVVEGKRVIMVDDSI VR 396

Query: 363 GTTSRRIVGLLREAGATEVHVVAIASPELKYPFCYGIDIQTRRELISANHAVDEVCDIIGA 422
GTTSRRIV LL++AGA EVHVAIASP LKYPFCYGIDIQ R ELI+A H DE+ + IGA

35 Sbjct: 397 GTTSRRIVKLLKDAGAAEVHVAIASPALKYPFCYGIDIQDRDELIAATHTTDEIREAIGA 456

Query: 423 DSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEEYLRSL 472
DSLTYLS GL+++IG + LC++YFDG YPTPLYDYE +YL SL

Sbjct: 457 DSLTYLSQSGLVEAIG-----HDKLCLSYFDGEYPTPLYDY EADYLESL 500

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 973> which encodes the amino acid sequence <SEQ ID 974>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0610(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 473/484 (97%), Positives = 481/484 (98%)

Query: 1 MTYEVKSLNEECGVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGL 60
MTYEVKSLNEECGVFGIWG+PQAAQVTYFGLHSLQHRGQEGAGI+SNDNGKLYGYRNVGL

55 Sbjct: 20 MTYEVKSLNEECGVFGIWGHPQAAQVTYFGLHSLQHRGQEGAGIVSNDNGKLYGYRNVGL 79

Query: 61 LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFYKFHDGQFALCHNGNLTNAI 120
LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFYKFHDGQFALCHNGNLTNAI

60 Sbjct: 80 LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFYKFHDGQFALCHNGNLTNAI 139

Query: 121 SSRKELEKQGAIFNASSDTEILMHLIRSHNPSFMGKVKKEALSTVKGGFAYLLMTEDKLI 180
S RKELEKQGAIFNASSDTEILMHLIRSHN SFMGKVKKEAL+TVKGGFAYLLMTE+KLI

Sbjct: 140 SLRKELEKQGAIFNASSDTEILMHLIRSHNSSFMGKVKKEALNTVKGGFAYLLMTENKLI 199

-386-

Query: 181 AALDPNAPRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYT 240
 AALDPNAPRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDD GIQCDRYT
 Sbjct: 200 AALDPNAPRPLSIGQMONGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDRGIQCDRYT 259

5 Query: 241 DETQLAICSMYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 300
 DETQLAICSMYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA
 Sbjct: 260 DETQLAICSMYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 319

10 Query: 301 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVGKRVVMIDDSI 360
 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVGKRVVMIDDSI
 Sbjct: 320 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVGKRVVMIDDSI 379

15 Query: 361 VRGTTSRRIVGLLREAGATEVHVAIASPELKYPFCFYGIDIQTRRELISANHAVDEVCDII 420
 VRGTTSRRIVGLLREAGA+EVHVAIASPELKYPFCFYGIDIQTRRELISANH+VDEVCDII
 Sbjct: 380 VRGTTSRRIVGLLREAGASEVHVAIASPELKYPFCFYGIDIQTRRELISANHSVDEVCDII 439

20 Query: 421 GADSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEYLRSLSEKTSFYI 480
 GADSLTYLS+DGLI+SIGLETKAPNGGLCVAYFDGHYPTPLYDYEYLRSLSEKTSFYI
 Sbjct: 440 GADSLTYLSLDGLIESIGLETKAPNGGLCVAYFDGHYPTPLYDYEYLRSLSEKTSFYI 499

Query: 481 QKVK 484
 QKVK
 Sbjct: 500 QKVK 503

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 302

A DNA sequence (GBSx0331) was identified in *S.agalactiae* <SEQ ID 975> which encodes the amino acid sequence <SEQ ID 976>. Analysis of this protein sequence reveals the following:

30 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.4797(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 303

A DNA sequence (GBSx0332) was identified in *S.agalactiae* <SEQ ID 977> which encodes the amino acid sequence <SEQ ID 978>. Analysis of this protein sequence reveals the following:

45 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3489(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-387-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 304

A DNA sequence (GBSx0333) was identified in *S.agalactiae* <SEQ ID 979> which encodes the amino acid sequence <SEQ ID 980>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.1690 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:CAC12194 GB:AL445066 phosphoribosylformylglycinamide synthase
 related protein [Thermoplasma acidophilum]
 Identities = 199/746 (26%), Positives = 329/746 (43%), Gaps = 103/746 (13%)

20 Query: 202 ADD--FAAYKAEQGLAMEVDDLFIQDYFKSIGRVPTETELKVLDTYWSDHCRHTTFETE 259
 ADD A GLA+ +D++ ++ YF+ +GR P + E+ + WS+HC + + +
 Sbjct: 11 ADDARLKAISKRLGLALSLEDEKAVRSYFERLGRDPIDAEIHAVAQSWSEHCSYKSSKYY 70

 Query: 260 LKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKFPQTLMDMATIFGRYERANGRLDDMEV 319
 LK K+ L+ Y +AM D+ G
25 Sbjct: 71 LK-----KYLGSCLKTDYT-ILAMEDDAG----- 92

 Query: 320 SDEINACSVIEVDVGVKEPWLMLFKNETHNHPTEIEPFGGAATCIGGAIRDPLSGRSY 379
 VD DG + K E+HNHP+ +EP+GGAAT IGG +RD L +
30 Sbjct: 93 -----VVDFDG---EYAVVLKMHSHNHPSAVEPYGGAATGIGGIVRDVLCMGAQ 138

 Query: 380 VYQAMRISGAGDITTPIAETRAGKLPQQVISKTAHGYSSYGNQIGLATTYVREYFHPGF 439
 + GD+++ E G L + I G YGN+IG+ YF +
 Sbjct: 139 PVALIDSLFLGDVSSDRYE---GLLSPRYIFGGVVGIRDYGNRIGIPNVAGSLYFDKLY 195

35 Query: 440 VAKRMELGAVVGAAPKENVVREKP-EAGDVVVLLGGKTGRDGVGGATGSSKVQTVESVET 498
 + + VG ++ +VR K + GDV+VL+GGKTGRDG+ G +S + ++
 Sbjct: 196 NSNPLVNAGCVGIVRRDRIVRSKSYKPGDVLVLMGGKTGRDGIHGVNFASTTLG-KVTKS 254

40 Query: 499 AGAEVQKGNAIEERKIQRLEFRDGNVTRLIKKSNDFGAGGVCVAIGELAD---GLEIDL 554
 + +Q GN I E+ + + + N LI+ D G GG+ A E+ G EI LD
 Sbjct: 255 SRLAIQLGNPIVEQPMIKAVLEANDAGLIRAMKDLGGGGLSSAATEMVYAGFGAEITLD 314

 Query: 555 KVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAFIAACNKENIDAVVVATVTEKPNLVM 614
 + LK ++G EI ISESQERM + P DV+ K N+D V+ VT + +
45 Sbjct: 315 DIKLKESNMMSGWEIWEISESQERMLMECPEDVEKIRQIAEKWNLDVSVIGQVTADRRIRV 374

 Query: 615 TWNGETIVDLERCFLDTNGV-RVVVDAKVVDKDLTVPEARITTSATLEADMLKVLSDLNH 673
 + I+D++ FLD + V + K V+K +TVP+ E L + + ++ LN
50 Sbjct: 375 YYKKRKIIDMDIEFLDDSPVYQRPYRIKEVEKSVTVFPQ----EPEDLNSFVRDFMARLNT 430

 Query: 674 ASQKGLQTIQDSSVGRSTVNHPIGGR-YQITPTESSVQKLPVQYGVTTTASVMAQGYNPF 732
 ++ + +D +V ST+ P GR + T +++V K P++ + V+ G P
 Sbjct: 431 CARFNVVRQYDHTVRGSTIVTPFVGRPNKETHADATVIK-PLENSM--RGLVLTSGSRPN 487

55 Query: 733 IAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQYFERMDKQAEFRGQPVSAALLGSI 792
 + PY G + EA +++TG R ++ E GQ V ++
 Sbjct: 488 MVSVDPYAGTLLTLAEAYKNILSTG---GRPHSVVDALNFGNPEREEMGQFVESVRAIG 544

60 Query: 793 EAQIQGLPSIGGKDSMSGTFEBELTVPPTLVAFGVTTADS-RKVLSPFEKAAGENIY--- 848
 + + GLP + G S + + + PT V D R+ + K +G IY
 Sbjct: 545 DFCRKMGLPVVAGNVSFYNEYRKTDMPTPTIMMVGILDDVRRSRTTYMKGSGNAYILIG 604

-388-

Query: 849 -----YIPGQAISEDIDFDLIKANF--SQFEAIQAQHKITAASAVKYGG 890
 Y G + D+D +F S+ + I + H +++ GG
 Sbjct: 605 EPCDNLTGSEYSRMHGYTDGFLPAPDLDELTRIRDFLSSKADMILSSHVSS-----GG 658

5 Query: 891 VLESIALMTFGNRIGASVEIAELDSS 916
 + +L+ M+FG+ IG V+I+ + ++
 Sbjct: 659 LFAALSEMSFGSGIGFHVDISNVSA 684

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 981> which encodes the amino acid
 10 sequence <SEQ ID 982>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1415(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 1219/1256 (97%), Positives = 1226/1256 (97%)

Query: 11 SSYFRVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLASLKDLRIVQVYDVF 70
 SSYF VAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQL SLK LRIVQVYDVF
 Sbjct: 2 SSYFPVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLTSLKALRIVQVYDVF 61

25 Query: 71 NLAEDLLARA EKHI FSEQVTDRLTETAEITAE LDKVAFFAIEALPGQFDQRAASSQEALL 130
 NLAEDLLARA EKHI FSEQVTD LLTE EITAE LDKVAFFAIEALPGQFDQRAASSQEALL
 Sbjct: 62 NLAEDLLARA EKHI FSEQVTDCLLTETETAE LDKVAFFAIEALPGQFDQRAASSQEALL 121

30 Query: 131 LLGSDSQVKVNTAQLYLVNKDIAEAELEAVKNYLLNPVDSRFKDITLPLEVQAFSVSDKT 190
 L GSDSQVKVNTAQLYLVNKDI EAELEAVKNYLLNPVDSRFKDITLPLE QAFSVSDKT
 Sbjct: 122 LFGSDSQVKVNTAQLYLVNKDITEAELEAVKNYLLNPVDSRFKDITLPLEEQAFSVSDKT 181

35 Query: 191 ISNLDFFETYQADDFAAAYKAEQGLAMEVDDLFIQDYFKSIGRVPTETELKVLDTYWSDH 250
 I NLDFFETYQADDF A YKAEQGLAMEVDDLFIQ+YFKSIG VPTETELKVLDTYWSDH
 Sbjct: 182 IPNLDFFETYQADDFATYKAEQGLAMEVDDLFIQNYFKSIGCVPTETELKVLDTYWSDH 241

40 Query: 251 CRHTTFETELKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA 310
 CRHTTFETELKNIDFSASKFQKQLQ TYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA
 Sbjct: 242 CRHTTFETELKNIDFSASKFQKQLQTTYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA 301

45 Query: 311 NGRLLDMEVSD E INACSVEIEVDVDGVKEPWLMLFKNETHNHPTIEIEPFGGAATCIGGAI 370
 NGRLLDMEVSD E INACSVEIEVDVDGVKEPWLMLFKNETHNHPTIEIEPFGGAATCIGGAI
 Sbjct: 302 NGRLLDMEVSD E INACSVEIEVDVDGVKEPWLMLFKNETHNHPTIEIEPFGGAATCIGGAI 361

50 Query: 371 RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTA AHGYSSYGNQIGLATTY 430
 RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTA AHGYSSYGNQIGLATTY
 Sbjct: 362 RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTA AHGYSSYGNQIGLATTY 421

55 Query: 431 VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVLLGGKTGRDGVGGATGSSKV 490
 VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVV+LLGGKTGRDGVGGATGSSKV
 Sbjct: 422 VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVILLGGKTGRDGVGGATGSSKV 481

60 Query: 491 QTVESVETAGAEVQKGNAIEERKIQR LFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 550
 QTVESVETAGAEVQKGNAIEERKIQR LFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE
 Sbjct: 482 QTVESVETAGAEVQKGNAIEERKIQR LFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 541

65 Query: 551 IDLDKVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAFIAACNKENIDAVVVATVTEKP 610
 IDLDKVPLKYQGLNGTEIAISESQERMSVVV P+DVDAFIAACNKENIDAVVVATVTEKP
 Sbjct: 542 IDLDKVPLKYQGLNGTEIAISESQERMSVVVRPNDVDAFIAACNKENIDAVVVATVTEKP 601

Query: 611 NLVMTWNGETIVDLERCFLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEADMLKVLSD 670
 NLVMTWNGE IVDLER FLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEAD LKVLSD
 Sbjct: 602 NLVMTWNGE IVDLERRFLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEADTLKVLSD 661

Query: 671 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQYGVTTTASVMAQQYN 730
 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQ+GVTTTASVMAQQYN
 Sbjct: 662 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQHGVTTTASVMAQQYN 721

5 Query: 731 PYIAEWSPHYGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG 790
 PYIAEWSPHYGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG
 Sbjct: 722 PYIAEWSPHYGAAYAVIEATARLVATGADWSRARFSYQYEFERMDKQAEFRGQPVSAALLG 781

10 Query: 791 SIEAQIQFGLPSIGGKDSMSGTFEELTVPPTLVAFGVTTADSRKVLSPPEKAAGENIYII 850
 SIEAQIQ GLPSIGGKDSMSGTFE+LTVPPTLVAFGVTTADSRKVLSPPEKAAGENIYII
 Sbjct: 782 SIEAQIQLGLPSIGGKDSMSGTFEDLTVPPTLVAFGVTTADSRKVLSPPEKAAGENIYII 841

15 Query: 851 PGQAISEDIDFDLIKANFSQFEAIQAQHKITAASAVKYGGVLESALMTFGNRIGASVEI 910
 PGQAISEDIDFDLIK NFSQFEAIQAQHKITAASA KYGGVLESALMTFGNRIGASVEI
 Sbjct: 842 PGQAISEDIDFDLIKDNFSQFEAIQAQHKITAASAAKYGGVLESALMTFGNRIGASVEI 901

20 Query: 911 AELDSSLTAQLGGFVFTSVEEADIADVVKIGQTQADFTVTVNGNDLAGASLLSAFEGKLEEV 970
 AELDSSLTAQLGGFVFTS EEIAD VKIGQTQADFTVTVNGNDLAGASLL+AFEGKLEEV
 Sbjct: 902 AELDSSLTAQLGGFVFTSAEEIADAVKIGQTQADFTVTVNGNDLAGASLLAAFEKLEEV 961

25 Query: 971 YPTEFEQVDAIEEVPVAVSDVVIKAKEIIEKPVVYIPVFFGTNSEYDSAKAFEQVGASVN 1030
 YPTEFEQ D +EEVPVAVSD VIKAKE IEKPVVYIPVFFGTNSEYDSAKAFEQVGASVN
 Sbjct: 962 YPTEFEQTDVLEEVPAVVSDTVIKAKETIEKPVVYIPVFFGTNSEYDSAKAFEQVGASVN 1021

30 Query: 1031 LVPFVTLNEAAIAESVDTMVANIAKANIIFAGGFSADEPDGSAKFIVNILLNEKVRAA 1090
 LVPFVTLNE AIAESVDTMVANIAKANIIFAGGFSADEPDGSAKFIVNILLNEKVRAA
 Sbjct: 1022 LVPFVTLNEVAIAESVDTMVANIAKANIIFAGGFSADEPDGSAKFIVNILLNEKVRAA 1081

35 Query: 1091 IDSPIEKGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1150
 IDSPIEKGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA
 Sbjct: 1082 IDSPIEKGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1141

40 Query: 1151 NTNSPFWLAGVEVGDIHVIPVSHGEGKFVVSASEFAELRDNGQIWSQYVDFDQPSMSDKY 1210
 NTNSPFWLAGVEVGDIH IPVSHGEGK VVSASEFAELRDNGQIWSQYVDFDQPSMSDKY
 Sbjct: 1142 NTNSPFWLAGVEVGDIHAIPVSHGEGKLVVSASEFAELRDNGQIWSQYVDFDQPSMSDKY 1201

Query: 1211 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIIPGNKDQKLFESAVKYFTGK 1266
 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIIPGNKDQ LF SAVKYFTGK
 Sbjct: 1202 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIIPGNKDQILFASAVKYFTGK 1257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 305

A DNA sequence (GBSx0334) was identified in *S. agalactiae* <SEQ ID 983> which encodes the amino acid sequence <SEQ ID 984>. This protein is predicted to be phosphoribosylaminoimidazole-succinocarboxamide synthase (purC). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4783(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA03540 GB:L15190 SAICAR synthetase [Streptococcus pneumoniae]
 Identities = 183/231 (79%), Positives = 203/231 (87%)

60 Query: 1 MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60
 M+ QLIY+GKAKDIY+T+DEN+I + YKDQAT NG +KE I GKG LNNQISS IFEKL
 Sbjct: 1 MSKQLIYSGKAKDIYTTEDENLIISTYKDQATAFNGVKKEQIAGKGVLLNNQISSFIFEKL 60

-390-

Query: 61 NMAGVVTHYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120
 N AGV TH++E++S EQLNKKV IIPLEVLRN TAGSFSKRFGV+EG LETPIVEFY
 Sbjct: 61 NAAGVATHFVEKLSDEQLNKKVKIIPLEVLRNVTAGSFSKRFGVDEGIALETPIVEFY 120

Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180
 YKND+L+DPFINDEHVKFL I +D++IAYLK E R INELLK WFA+IGL LIDFKLEFG
 Sbjct: 121 YKNDLDDPFINDEHVKFLQIADDQQAAYLKKEARRINELLKVWFAEIGLKLIDFKLEFG 180

Query: 181 FDKDGKIIILADEFSPDNCRLWDADGNHMDKDVFRRLGSLTDVYQVVLEKL 231
 FDKDGKIIILADEFSPDNCRLWDADGNHMDKDVFRRL LG LTDVY++V EKL
 Sbjct: 181 FDKDGKIIILADEFSPDNCRLWDADGNHMDKDVFRRLGSLTDVYVIVWEKL 231

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 985> which encodes the amino acid
 sequence <SEQ ID 986>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 221/234 (94%), Positives = 228/234 (96%)

Query: 1 MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60
 +TNQLIY GKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL
 Sbjct: 11 VTNQLIYKGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 70

Query: 61 NMAGVVTHYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120
 N AGVVTHYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY
 Sbjct: 71 NKAGVVTHYIEQISKNEQLNKKVDIIPLEVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 130

Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180
 YKND+L+DPFINDEHVKFLGIVNDEEIAYLKGETR INELLK WFAQIGLNLIDFKLEFG
 Sbjct: 131 YKNDLDDPFINDEHVKFLGIVNDEEIAYLKGETRRINELLKGWFAQIGLNLIDFKLEFG 190

Query: 181 FDKDGKIIILADEFSPDNCRLWDADGNHMDKDVFRRLGSLTDVYQVVLEKLIAL 234
 FD++G IILADEFSPDNCRLWD +GNHMDKDVFRRLG+LTDVYQVVLEKLIAL
 Sbjct: 191 FDQEGTIIILADEFSPDNCRLWDKNGNHMDKDVFRRLGSLTDVYQVVLEKLIAL 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 306

A DNA sequence (GBSx0335) was identified in *S.agalactiae* <SEQ ID 987> which encodes the amino acid
 sequence <SEQ ID 988>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9457> which encodes amino acid sequence <SEQ ID 9458>
 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-391-

>GP:AAC35700 GB:AF041468 acyl carrier protein [Guillardia theta]
Identities = 27/75 (36%), Positives = 52/75 (69%)

5 Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAI DSIALTEFIINLEDVFHLEIPDE 71
M+ E+FEK+ ++ +QLG + +T +++ +DL DS+ E ++ +E+ F++EIPD+
Sbjct: 1 MNEQEIFEKVKQTIISEQLGVDKSQVTKDANFANDLGADSLDTVELVMAIEEAFNIEIPDD 60

Query: 72 AVEHMSSVQQLLDYI 86
A E +S++QQ +D+I

10 Sbjct: 61 AAEQISNLQQA VDFI 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 989> which encodes the amino acid sequence <SEQ ID 990>. Analysis of this protein sequence reveals the following:

Possible site: 24
15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1917(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 36/77 (46%), Positives = 57/77 (73%)

25 Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAI DSIALTEFIINLEDVFHLEIPDE 71
M+R E+FE+++ L+++Q + IT ++ L +DLA+DSI L EFIIN+ED FH+ IPDE
Sbjct: 1 MTRQEIFERLINLIQKQRSYLSVAITEQTHLKNDLAVDSIELVEFIINVEDEFHIAIPDE 60

Query: 72 AVEHMSSVQQLLDYIIE 88
VE M ++ +LDY+++

30 Sbjct: 61 DVEDMVFMRDILDYLVQ 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 307

A DNA sequence (GBSx0336) was identified in *S.agalactiae* <SEQ ID 991> which encodes the amino acid sequence <SEQ ID 992>. This protein is predicted to be fatty acid/phospholipid synthesis protein (plsX). Analysis of this protein sequence reveals the following:

Possible site: 21
40 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.64 Transmembrane 101 - 117 (101 - 117)

----- Final Results -----
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9455> which encodes amino acid sequence <SEQ ID 9456> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13462 GB:Z99112 alternate gene name: ylpD [Bacillus subtilis]
Identities = 174/329 (52%), Positives = 238/329 (71%), Gaps = 2/329 (0%)

55 Query: 8 KIAIDAMGGDYAPKAIVEGVNQAISDFSDIEVQLYGDQKKIEKYLTVT-ERVSIIHTEEK 66
+IA+DAMGGD+APKA+++GV + I F D+ + L GD+ IE +LT T +R++++H +E
Sbjct: 2 RIAVDAMGGDHAPKAVIDGVIKGIEAFDDLHITLVGDKTTTIESHLTTTSDRITVLHADEV 61

-392-

Query: 67 INSDDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDRP 126
 I DEF +AVRRKK SSMVL A+ V + A A ISAGNTGAL+ AGLF+VGRIKG+DRP
 Sbjct: 62 IEPTDEPVRAVRRKKNSMVLMAQEVAENRADACISAGNTGALMTAGLFIVGRIKGIDRP 121
 Query: 127 GLMSTMPITLDGVGFDMLDLGANAENTASHLHQYAILGSAFYAKNVRGIEVPRVGLLNNGTE 186
 L T+PT+ G GF +LD+GAN + HL QYAI+GS Y++ VRG+ PRVGLLN GTE
 Sbjct: 122 ALAPTLETVSGDGFLLLDVGANVDAKPEHLVQYAIMGSVYSQQVRGVTSPRVGLLNNGTE 181
 Query: 187 ETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVTGFTGNAVLKTMEGTAMS 246
 + KG+ L K+ +++L +INFIGN+EARDL+ VADVVTGFTGN LKT+EG+A+S
 Sbjct: 182 DKKGNELTKQTFQILKETANINFIGNVEARDLDDVADVVTGFTGNVTLKTLGEGSALS 241
 Query: 247 IMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSDS 306
 I ++ + + +KL A +LK L ++K M+YS+ GGA LFGLKAP++K HGSSDS
 Sbjct: 242 IFKMMR-DVMTSTLTSLKLA AAVLKP KLEMMKMEYSNYGGASLFGLKAPVIKAHGSSDS 300
 Query: 307 KAVYSTLKQVRTMLETQVVDQLVDAFTDE 335
 AV+ ++Q R M+ V + + +E
 Sbjct: 301 NAVFHAIRQAREMVSQNVAALIQEEVKKE 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 993> which encodes the amino acid sequence <SEQ ID 994>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 121 - 137 (120 - 138)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9127> which encodes the amino acid sequence <SEQ ID 9128>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 16
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 95 - 111 (94 - 112)
 ----- Final Results -----
 bacterial membrane --- Certainty= 0.183(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 254/330 (76%), Positives = 290/330 (86%)
 Query: 6 MKKIAIDAMGGDYAPKAIVEGVNQAI SDFSDIEVQLYGDQKKIEKYLT TVERVSI IHTTE 65
 MK+IAIDAMGGD APKAIVEGVNQAI FSDIE+QLYGDQ KI YL ++RV+IIHT+E
 Sbjct: 27 MKRIAIDAMGGDNAPKAIVEGVNQAI EAFSDIEIQLYGDQTKINSYLIQSDRVAI IHTDE 86
 Query: 66 KINSDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDR 125
 KI SDDEPAKAVRRKK++SMVL AKAVK+G A A ISAGNTGALLA GLFVVGRIKGVDR
 Sbjct: 87 KINSDEPAKAVRRKKKASMVLA AKAVKEGKADAIISAGNTGALLAVGLFVVGRIKGVDR 146
 Query: 126 PGLMSTMPITLDGVGFDMLDLGANAENTASHLHQYAILGSAFYAKNVRGIEVPRVGLLNNGT 185
 PGL+ST+PT+ G+GFDMLDLGANAENTA HLHQYAILGSAFYAKNVRGI PRVGLLNNGT
 Sbjct: 147 PGLLSTIPTVTGLGFDMLDLGANAENTAKHLHQYAILGSAFYAKNVRGIANPRVGLLNNGT 206
 Query: 186 EETKGDLSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVTGFTGNAVLKTMEGTAM 245
 EETKGD L K YELL A+ +I+F+GN+EAR+LMS VADV+V+DGFTGNAVLK++EGTA+
 Sbjct: 207 EETKGDPLRKATYELLTADNTISFVGNVEARELMSGVADVIVSDGFTGNAVLKSI EGTAI 266
 Query: 246 SIMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSD 305

-393-

SIMG LK I SGG+K K+GA LLK SLY++K ++DYSSAGGAVLFGLKAP+VK HGSSD
 Sbjct: 267 SIMGQLKQIINSGGIKTKIGASLLKSSLYEMKKTLDYSSAGGAVLFGLKAPVVKSHGSSD 326

Query: 306 SKAVYSTLKQVRTMLETOVVDQLVDAFTDE 335
 KA++ST+KQVRTML+T VV QLV+ F E
 Sbjct: 327 VKAIFSTIKQVRTMLDTINVVGQLVEEFAKE 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 308

A DNA sequence (GBSx0337) was identified in *S.agalactiae* <SEQ ID 995> which encodes the amino acid sequence <SEQ ID 996>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4668 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 309

A DNA sequence (GBSx0338) was identified in *S.agalactiae* <SEQ ID 997> which encodes the amino acid sequence <SEQ ID 998>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.84 Transmembrane 61 - 77 (55 - 82)
 INTEGRAL Likelihood = -10.14 Transmembrane 26 - 42 (19 - 51)
 INTEGRAL Likelihood = -9.77 Transmembrane 192 - 208 (186 - 211)
 INTEGRAL Likelihood = -5.79 Transmembrane 267 - 283 (262 - 286)
 INTEGRAL Likelihood = -3.77 Transmembrane 100 - 116 (99 - 116)

----- Final Results -----
 bacterial membrane --- Certainty=0.6137 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9453> which encodes amino acid sequence <SEQ ID 9454> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA22372 GB:AL034446 putative transmembrane protein
 [Streptomyces coelicolor A3(2)]
 Identities = 47/154 (30%), Positives = 69/154 (44%), Gaps = 12/154 (7%)

Query: 120 SGFVEISSNSFSFGPPFFFLFLAYFIQSLTEEILFRGYVMTIVTKFGSFAGVLCNSMLF 179
 SG+ E+ S F+A + TEE++FRG + + + G++ + ++F
 Sbjct: 118 SGYYEVDGLGSGVQGAIGLVGFMA--AAAATEEVVFRGVLFRIIEEHIGTYLALGLTGLVF 175

Query: 180 SFIHFRN----YGITAIALFNLFLGLGIIFSILFNMTKNILFVTGVHTTWNFTMGCVLGN 234
 +H N +G AIA+ F+L ++ T+N+ GVH WNF G V

-394-

Sbjct: 176 GLMHLLNEDATLWGALAIATEAGFMLAAAYAA----TRNLWLTIGVHFGWNFAAGGVFST 231

Query: 235 KVSQGDSPVSLFRITENSSFALWNGDGFEGGV 268
VSG L T S L GDFG EG V

5 Sbjct: 232 VVSGNGDSEGLLDAT-MSGPKLLTGDFGPEGSV 264

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 310

A DNA sequence (GBSx0339) was identified in *S.agalactiae* <SEQ ID 999> which encodes the amino acid sequence <SEQ ID 1000>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2665(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9451> which encodes amino acid sequence <SEQ ID 9452> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05088 GB:AP001511 unknown conserved protein [Bacillus halodurans]
Identities = 81/242 (33%), Positives = 124/242 (50%), Gaps = 3/242 (1%)

25

Query: 8 GLVLYNRNYREDDKLVKIFTETEGKRMFFVKHAS--KSKFNAVLQPLTIAHFILKINDNG 65
G+V+ +Y E +K+V +FT GK + A KS+ AV Q T + + N G

30

Sbjct: 7 GIVIRTVDYGESNKIVTVFTREYGKIALMARGAKRPKSRLTAVTQLFTYGMFMFQKNA-G 65

Query: 66 LSYIDDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTILELIEDG 125
L + + + +F+E +DLF+ SY SY+T L + D + LF L +T+ + +G

Sbjct: 66 LGTLTQGEIIQSFREVRNDLFRASYVSVYTDLTNKLTEDEKRNPFYLFELLYQTIHYMNEG 125

35

Query: 126 LDYEILTNIFEVQLLERFGVALNFHDCVFCFHRVGLPFDPSHKYSGLLCPNHYYKDERRNH 185
+D ++LT IFEV++ G+ CV C +P FS K +G LC KD

Sbjct: 126 MDPDLVLRIFEVKMFTVAGIKPELDQCVCSTDPVPGFSIKEAGFLCKRCIEKDPHAYK 185

40

Query: 186 LDPNMLYLINRFQSIQFDDLTQTSVKPEMKLKIRQFLDMIYDEYVGIHLKSKKFIDDLSSWG 247
+ + L+ F L TIS+KPE K ++ + YDEY G+HLKS++F+D L S G

Sbjct: 186 ITAQVAKLLRLFYHFDLQRLGTISLKPETKATLKTIIHQYYDEYSGHLKSRFLDQLESWG 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1001> which encodes the amino acid sequence <SEQ ID 1002>. Analysis of this protein sequence reveals the following:

45

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1566(Affirmative) < succ>

50

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 159/251 (63%), Positives = 210/251 (83%)

55

Query: 1 MRVSQTYGLVLYNRNYREDDKLVKIFTETEGKRMFFVKHASKSKFNAVLQPLTIAHFILK 60

-395-

M+++++ G+VL+NRNYREDDKLVKIFTE GK+MFFVKH S+SK ++++QPLTIA FI K
 Sbjct: 1 MQLTESLGIVLFNRNYREDDKLVKIFTEVAGKQMFVVKHISRSMSSIIQPLTIADFIK 60
 Query: 61 INDNGLSYIDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTL 120
 +ND GLSY+ DY V ++ N+D+F+L+YASY+ +LAD AI+DN +D+ LF FLKKTL+
 Sbjct: 61 LNDTGLSYVVDYSNVNTYRYINNDIFRLAYASYVLALADAAIADNESDHLFTFLKKTL 120
 Query: 121 LIEDGLDYEILTNI FEFVQLLERFGVALNFHDCV FCHRVGLPFD FSHKYSGLLCPNHYYKD 180
 L+E+GLDYEILTNI FE+Q+L+RFG++LNFH+C CHR LP DFSH++S +LC HYYKD
 Sbjct: 121 LMEEGLDYEILTNI FEIQILDRFGISLNFHECAICHR TDLP LDFSHRFSAVLCSEHYYKD 180
 Query: 181 ERRNHLDPNMLYLINRFQSIQFDDLQ TISVKPEMKL KIRQFLDMIYDEYVGIHLKSKKFI 240
 RRNHLDPN++YL++RFQ I FDDL+TIS+ ++K K+RQF+D +Y +YVGI LKSK FI
 Sbjct: 181 NRRNHLDPNVYLLSRFQKITFDDLRTISLNKDIKKLRQFIDELYHDYVGIKLSKTFI 240
 Query: 241 DDLSSWGSIMK 251
 D+L WG IMK
 Sbjct: 241 DNLVKWGDIMK 251

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 311

A DNA sequence (GBSx0340) was identified in *S. agalactiae* <SEQ ID 1003> which encodes the amino acid sequence <SEQ ID 1004>. This protein is predicted to be aromatic amino acid aminotransferase (patA). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.13 Transmembrane 141 - 157 (140 - 159)

----- Final Results -----
 bacterial membrane --- Certainty=0.2253 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9449> which encodes amino acid sequence <SEQ ID 9450> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF06954 GB:AF146529 aromatic amino acid aminotransferase
 [Lactococcus lactis subsp. cremoris]
 Identities = 261/391 (66%), Positives = 323/391 (81%)
 Query: 38 MTLEKRFNKYLDRIEVS LIRQFDQSISDIPGMVKLT LGEPDFTT PDHVKEA AAKSAIDANQ 97
 M L K+FN LD+IE+SLIRQFDQ +S IP ++KLT LGEPDF TP+HVK+A +AI+ NQ
 Sbjct: 1 MDLLKKFNPNLDKIEIS LIRQFDQVSSIPDI IKLT LGEPDFYTP EHVQAGIAAIENNQ 60
 Query: 98 SYYTGM SGLLALRQAAADFAKDKYNLTYPDCEILVTIGATEALSASLIAILEAGDVVLL 157
 S+YTGM+GLL LRQAA++F KY L+Y + EILVT+G TEA+S+ L++IL AGD VL+
 Sbjct: 61 SHYTGMAGLLELRQAA SEFLKKYGLSYAAEDELVTIVGVTEA ISSVLLSILVAGDEVLI 120
 Query: 158 PAPAYPGYEPIVNLVGADIVEIDTREND FRLTP EMLETAI IQQGEKLKAVLLNYP TNPTG 217
 PAPAYPGYEP++ L G +VEIDTR NDF LTP EML+ AII++ K+KAV+LNYP NPTG
 Sbjct: 121 PAPAYPGYEPLITLAGGSLVEIDTRANDFVLTPEMLDQAI IEREGKV KAVILNYPANPTG 180
 Query: 218 ITYSRQEIAALAEVLKKYDIFVISDEVYSELTYTQQQHVSIAEYLPNQITILINGLSKSHA 277
 +TYR++I LAEVLKK+++FVI+DEVYSEL YT Q HVSIAEY P QTI++NGLSKSHA
 Sbjct: 181 VTYNREQIKDLAEVLKKHEVFVIADEVYSELNYTDQPHVSIAEYAPEQTITVLNGLSKSHA 240
 Query: 278 MTGWRVGLVYAPEAFIAQIIKSHQYMTAASTISQFAGVEALSVGKNDTLPMRQGYIKRR 337
 MTGWR+GL++A +AQIIK+HQY+VT+AST SQFA +EAL G +D LPM++ Y+KRR

-396-

Sbjct: 241 MTGWRIGLIFAARELVAQIIKTHQYLVTSASTQSQFAAIBALKNGADDALPMKEYLKRR 300

Query: 338 DYIIDKMSKLGFKIIPSGAFYIFAKIPDSYPQDSFKFCQDFAYQQAVAIIPGVAFGKYG 397
DYII+KMS LGFKII+P GAFYIFAKIP QDSFKF DFA + AVAIIPG+AFG+YG

5 Sbjct: 301 DYIIEKMSALGFKIIEPDGAFYIFAKIPADLEQDSFKFAVDFAKENAVAIIPGIAFGQYG 360

Query: 398 EGYIRLSYAASMEVIEETAMARLKVFMESEYEG 428
EG++RLSYAASM+VIE AMARL ++ G

10 Sbjct: 361 EGFVRLSYAASMDVIEQAMARLTDYVTKKRG 391

There is also homology to SEQ ID 1006.

SEQ ID 1004 (GBS332) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 3; MW 50.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 4; MW 76kDa).

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 312

A DNA sequence (GBSx0341) was identified in *S.agalactiae* <SEQ ID 1007> which encodes the amino acid sequence <SEQ ID 1008>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA).

20 Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3118(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9447> which encodes amino acid sequence <SEQ ID 9448> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62181 GB:M92842 prs [Listeria monocytogenes]
Identities = 209/312 (66%), Positives = 266/312 (84%), Gaps = 3/312 (0%)

35 Query: 10 LKLFALSSNKELAKKVSQTIGIPLGQSTVRQFSDGEIQVNIIEESIRGHVFIHQSTSSPV 69
LK+F+L+SN+ELA++++ +GI LG+S+V FSDGEIQ+NIIEESIRG HV+++QSTS+PV
Sbjct: 10 LKIFSLNSNRELAEEIAKEVGIELGKSSVTHPSDGEIQINIEESIRGCHVYVIQSTSNPV 69

40 Query: 70 NDNLMELIMVDALKRASAESVSVMPYGYARQDRKARSREPITSKLVANMLEVAGVDR 129
N NLME+LIM+DALKRASA ++++VMPYGYARQDRKARSREPIT+KLVAN++E AG R
Sbjct: 70 NQNLMEELIMIDALKRASAATINIVMPYGYARQDRKARSREPITAKLVANLIETAGATR 129

45 Query: 130 LLTVDLHAAQIQGFDDIPVDHLMGAPLIADYFDRQGLVGDVVVSPDHGGVTRARKLAQ 189
++T+D+HA QIQGFDDIP+DHL L++DYF + L GDD+VVVSPDHGGVTRARK+A
Sbjct: 130 MITLDMHAPQIQGFDDIPIDHLNAVRLLSDYFSEHRL-GDDLVVVSPDHGGVTRARKMAD 188

50 Query: 190 CLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHAADALAEAGA 249
LK PIAIIDKRR + N +EVMNI+GN++GK CI+IDD+IDTAGTI AA AL EAGA
Sbjct: 189 RLKAPIAIIIDKRR--PRPNVAEVMNIVGNVEGKVCIIIDDIIDTAGTITLAALKALREAGA 246

55 Query: 250 TAVYASCTHPVLSGPALDNQNSAIEKLIVLDTIYLPPEERLIDKIEQISIAELIGEAIIR 309
T VYA C+HPVLSGPA+ I+ S IEKL+V ++I LPPE+ IDK+EQ+S+A L+GEAI+R
Sbjct: 247 TKVYACCSHPVLSGPAMKRIEESPIELVVTNSIALPEEKWIDKMEQLSVAALLGEAIVR 306

Query: 310 IHEKRPLSPLFE 321
+HE +S LFE

-397-

Sbjct: 307 VHENASVSSLFE 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1009> which encodes the amino acid sequence <SEQ ID 1010>. Analysis of this protein sequence reveals the following:

5 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

15 Identities = 298/322 (92%), Positives = 311/322 (96%)

Query: 1 MEEIMSYSNLKLFALSSNKELAKKVSQTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVF 60
 +EE MSYS+LKLALSSNKELA+KV+ +GI LG+STVRQFSDGEIQVNIEESIRGHHVF
 Sbjct: 1 LEEKMSYSDLKLFALSSNKELAEKVASAMGIQLGKSTVRQFSDGEIQVNIEESIRGHHVF 60

20 Query: 61 ILQSTSSPVNDNLMEILIMVDALKRASAESVSVMPYYGYARQDRKARSREPITSKLVAN 120
 ILQSTSSPVNDNLMEILIMVDALKRASAE +SVMPYYGYARQDRKARSREPITSKLVAN
 Sbjct: 61 ILQSTSSPVNDNLMEILIMVDALKRASA EKISVMPYYGYARQDRKARSREPITSKLVAN 120

25 Query: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDROGLVGDDVVVSPDHGG 180
 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDR GLVG+DVVVVSPDHGG
 Sbjct: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRHGLVGEDVVVSPDHGG 180

30 Query: 181 VTRARKLAQCLKTPIAIIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHA 240
 VTRARKLAQ L+TPIAIIIDKRRSV KMNTSEVMNIIGN+ GKKCILIDDMIDTAGTICHA
 Sbjct: 181 VTRARKLAQFLQTPIAIIIDKRRSVDKMNTSEVMNIIGNVSGKKCILIDDMIDTAGTICHA 240

35 Query: 241 ADALAEAGATAVYASCTHPVLSGPAIDNIQNSAIEKLIVLDTIYLP+ERLIDKIEQISIA 300
 ADALAEAGATAVYASCTHPVLSGPAIDNIQ SAIEKLIVLDTIYLP+ERLIDKIEQISIA
 Sbjct: 241 ADALAEAGATAVYASCTHPVLSGPAIDNIQNSAIEKLIVLDTIYLPKERLIDKIEQISIA 300

40 Query: 301 ELIGEAIIRIHEKRPLSPLFEM 322
 +L+ EAIIRIHEKRPLSPLFEM
 Sbjct: 301 DLVAEAIIRIHEKRPLSPLFEM 322

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 313

A DNA sequence (GBSx0342) was identified in *S.agalactiae* <SEQ ID 1011> which encodes the amino acid sequence <SEQ ID 1012>. This protein is predicted to be a secreted protein. Analysis of this protein
 45 sequence reveals the following:

 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9277> which encodes amino acid sequence <SEQ ID 9278> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

30 The GBS65-GST fusion product was purified (Figure 102A; see also Figure 191, lane 4) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 102B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 314

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1184(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-400-

Example 315

A DNA sequence (GBSx0344) was identified in *S.agalactiae* <SEQ ID 1017> which encodes the amino acid sequence <SEQ ID 1018>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.4736(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
15 vaccines or diagnostics.

Example 316

A DNA sequence (GBSx0345) was identified in *S.agalactiae* <SEQ ID 1019> which encodes the amino acid sequence <SEQ ID 1020>. This protein is predicted to be elongation factor Tu (tufA). Analysis of this protein sequence reveals the following:

```

20  Possible site: 43
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3012(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9737> which encodes amino acid sequence <SEQ ID 9738> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB03851 GB:AP001507 translation elongation factor Tu (EF-Tu)
[Bacillus halodurans]
Identities = 302/397 (76%), Positives = 350/397 (88%), Gaps = 2/397 (0%)

35  Query: 7  MAKEKYDRSKPHVNIQTIGHVDHGKTTTLTAAITTVLARRLPTSVNQPKDYASIDAAPPEER 66
      MAKEK+DRSK H NIGTIGHVDHGKTTTLTAAITTVLA+R V Y +ID APEER
      Sbjct: 1  MAKEKFDRSKTHANIGTIGHVDHGKTTTLTAAITTVLAKRSGKGVAMA--YDAIDGAPPEER 58

      Query: 67  ERGITINTAHVEYETEKRYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDGPMPQTR 126
      ERGITI+TAHVEYET+ RHYAH+D PGHADYVKNMITGAAQMDG ILVV++ DGPMPQTR
40  Sbjct: 59  ERGITISTAHVEYETDNRHYAHVDCPGHADYVKNMITGAAQMDGGILVVSADGMPMPQTR 118

      Query: 127  EHILLSRQVGKHLIVFMNKKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIVQGSALK 186
      EHILLSRQGVG +L+VF+NK D+VDDEELLELVEME+RDLLSEYDFPGDD+PVI+GSALK
45  Sbjct: 119  EHILLSRQGVGPYLVFLNKCMDVDDEELLELVEMEVRDLLSEYDFPGDDVPVIRGSALK 178

      Query: 187  ALEGDEKYEDIIMELMSTVDEYIPEPERDTPKPLLLFVEDVFSITGRGTVASGRIDRGTV 246
      ALEGD ++E+ I+ELM+ VD+YIP PERDT+KP ++PVEDVFSITGRGTVA+GR++RG +
      Sbjct: 179  ALEGDAEWEEKIIELMMAVDDYIPTPERDTEKPFMFVVEDVFSITGRGTVATGRVERGQL 238

50  Query: 247  RVNDEVEIVGIKEDIQKAVVTGVEMFRKQLEGLAGDNVGVLLRGVQRDEIERGOVLAKP 306
      V DEVEI+G++E+ +K VTGVEMFRK LD AGDN+G LLRGV R+E++RGQVLAKP
      Sbjct: 239  NVGDEVEIIGLEEEAKKTTVTGVEMFRKLLDYAEAGDNIGALLRGVSRREEVQRGOVLAKP 298

```

-401-

Query: 307 GSINPHTRFKGEVYILSKEEGGRHTPFFNNYRQFYFRITTDVTGSIELPAGTEMVMPGDN 366
 G+I PHT FK EVY+LSKEEGGRHTPFF+NYRQFYFRITTDVTG I+LP G EMVMPGDN
 Sbjct: 299 GTITPHTNFKAEEVYVLSKEEGGRHTPFFSNYRQFYFRITTDVTGIIQLPDGVEMVMPGDN 358

5 Query: 367 VTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIE 403
 V + VELI PIA+E+GT FSIREGGRTVG+G+V+ I+
 Sbjct: 359 VEMTVELIPIAIEEGTKFSIREGGRTVGAGVVASIQ 395

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1021> which encodes the amino acid
 10 sequence <SEQ ID 1022>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.1367(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 386/404 (95%), Positives = 396/404 (97%)
 Query: 1 MEAFPKMAKEKYDRSKPHVNIQTIGHVDHGKTTLTAAITTVLARRLPTSVNQPKDYASID 60
 +EAFPKMAKEKYDRSKPHVNIQTIGHVDHGKTTLTAAITTVLARRLP+SVNQPKDYASID
 Sbjct: 12 LEAFPKMAKEKYDRSKPHVNIQTIGHVDHGKTTLTAAITTVLARRLPSSVNQPKDYASID 71
 25 Query: 61 AAPEERERGITINTAHVEYETEKRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 120
 AAPEERERGITINTAHVEYET RHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG
 Sbjct: 72 AAPEERERGITINTAHVEYETATRHAYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 131
 30 Query: 121 PMPQTRHILLSRQVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIVI 180
 PMPQTRHILLSRQVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIVI
 Sbjct: 132 PMPQTRHILLSRQVGKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPIVI 191
 35 Query: 181 QGSALKALEGDEKYEDIIMELMSTVDEYIPEPERDTPKPLLLPVEDVFSITGRGTVASGR 240
 QGSALKALEGD K+EDIIMELM TVD YIPEPERDTPKPLLLPVEDVFSITGRGTVASGR
 Sbjct: 192 QGSALKALEGDTKFEDIIMELMDTVDSYIPEPERDTPKPLLLPVEDVFSITGRGTVASGR 251
 Query: 241 IDRGTVRVNDEVEIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGLLRGVQRDEIERG 300
 IDRGTVRVNDE+EIVGIKE+ +KAVVTGVEMFRKQLDEGLAGDNVG+LLRGVQRDEIERG
 40 Sbjct: 252 IDRGTVRVNDEIEIVGIKEETKAVVTGVEMFRKQLDEGLAGDNVGILLRGVQRDEIERG 311
 Query: 301 QVLAKPGSINPHTRFKGEVYILSKEEGGRHTPFFNNYRQFYFRITTDVTGSIELPAGTEM 360
 QV+AKP SINPHT+FKGEVYILSK+EGGRHTPFFNNYRQFYFRITTDVTGSIELPAGTEM
 Sbjct: 312 QVIAPSSINPHTKFKGEVYILSKDEGGRTHTPFFNNYRQFYFRITTDVTGSIELPAGTEM 371
 45 Query: 361 VMPGDNVTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 404
 VMPGDNVTI VELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA
 Sbjct: 372 VMPGDNVTINVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 415

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 317

A DNA sequence (GBSx0346) was identified in *S.galactiae* <SEQ ID 1023> which encodes the amino
 acid sequence <SEQ ID 1024>. Analysis of this protein sequence reveals the following:

55 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -0.64 Transmembrane 90 - 106 (90 - 106)
 ----- Final Results -----
 60 bacterial membrane --- Certainty=0.1256(Affirmative) < succ>

-402-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 318

- 10 A DNA sequence (GBSx0347) was identified in *S.agalactiae* <SEQ ID 1025> which encodes the amino acid sequence <SEQ ID 1026>. This protein is predicted to be ftsW. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

- 15 INTEGRAL Likelihood = -11.15 Transmembrane 44 - 60 (35 - 70)
 INTEGRAL Likelihood = -4.73 Transmembrane 76 - 92 (74 - 98)
 INTEGRAL Likelihood = -3.88 Transmembrane 117 - 133 (113 - 134)

----- Final Results -----

- 20 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 25 >GP:AAE39929 GB:U58049 putative cell division protein ftsW
 [Enterococcus hirae]
 Identities = 78/159 (49%), Positives = 107/159 (67%), Gaps = 4/159 (2%)
- Query: 1 MANSXYAMSNGGWFGRLGNSIEKLGVLPEATTDVFVSIVIEELGVICAGFILALVFFLI 60
 M+NS YA+ NGG FGRG+GNSI K GYLPE+ TDF+FS++ EE G+IGA +L L+F L
 30 Sbjct: 240 MSNSYYALYNGGLFGRGMGNSITKGYLPESETDFIFSVIAEEFGLICALLVLFLLFLLC 299
- Query: 61 LRIMHVGIIKADPFNSMIALGIGAMLLMQVFNIGGISGLIPSTGVTFPFLSQGGNSLLV 120
 +RI K K+ ++I +G+G +L+Q +NIG I GLIP TGV PF+S GG S L+
 35 Sbjct: 300 MRIFQKSTKQKNQANLILIGVGTWILVQTSINIGSILGLIPMTGVPLPFVSYGGSYLI 359
- Query: 121 LSAIGFVLINIDANEKKELIMKEAEEQYKPEKNEKIIN 159
 LS AIG LNI + + KE + + + Q K K++N
 Sbjct: 360 LSFAIGLALNISRQVKE----KNKQVERLQLKPKLLN 394

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1027> which encodes the amino acid sequence <SEQ ID 1028>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

- 45 INTEGRAL Likelihood = -10.93 Transmembrane 312 - 328 (303 - 338)
 INTEGRAL Likelihood = -8.23 Transmembrane 22 - 38 (17 - 47)
 INTEGRAL Likelihood = -6.85 Transmembrane 192 - 208 (187 - 211)
 INTEGRAL Likelihood = -5.10 Transmembrane 218 - 234 (212 - 236)
 INTEGRAL Likelihood = -4.83 Transmembrane 86 - 102 (85 - 107)
 INTEGRAL Likelihood = -3.72 Transmembrane 385 - 401 (383 - 402)
 50 INTEGRAL Likelihood = -3.45 Transmembrane 61 - 77 (61 - 79)
 INTEGRAL Likelihood = -2.39 Transmembrane 344 - 360 (344 - 360)

----- Final Results -----

- 55 bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB59721 GB:AJ250603 FtsW protein [Enterococcus faecium]

Identities = 131/397 (32%), Positives = 223/397 (55%), Gaps = 23/397 (5%)

5 Query: 15 KRHLNYSILLPYLILSVIGLIMVYSTTSVSLIQAHANPFKSVINQGVFWIISLVAITFI 74
KR +++ IL PYL LS+IGL+ VYS +S L+QA N ++ Q +F +S I
Sbjct: 3 KRKKIDWWILGPYLTLISMIGLLEVYSASSYRLQADENTKSLLLRQLIFIFLSWGVIFLA 62

10 Query: 75 YKLKLNFLTNTRVLTVMVLGEAFLLIAR--FFTAKGAHGWIWIGPVSFQPAEYLKII 132
+KL++L + ++ + F LI+ R F + GA WI + + FQP+E +
Sbjct: 63 RSIKLHYLLHPKIAGYGLALSIFLILVRVGIFGVTVNGAQRWISLFGIQFQPSLANLF 122

15 Query: 133 MVWYLALTFAKIQKNISLYDYQALTRRKWWPTQWNDLRDWRVYSLLMVLLVAAQPDLGNA 192
+++YL+ F P + +L+ + ++ + LL+ QP + A
Sbjct: 123 LIFYLSWFFRDGNN-----PKK--NLKKPFLITVSITLLILFQPKIAGA 164

20 Query: 193 SIIVLTAIMFSISGIGYRWFSAIIVMITGLSTVFLGTIAVIGVERVAKIP-VFGYVAKR 251
+I+ A ++F + + ++ ++V + L G + +G + +P +F + +R
Sbjct: 165 LMILSIAWVIFWAAAVPFKKGIYLVTFSAIILIGAAGGVLYLGNK--GWLPMQMFNHAYER 222

25 Query: 252 FSAFFNPFHDLTDSGHQLANSYYAMSGGWFGQGLGNSIEKRGYLPEAQTDFVFSVVEE 311
+ +PF D +G+Q+ +S+YA+ NGG +G+GLGNSI K+GYLPE +TDF+FS++ EE
Sbjct: 223 IATLRDPFIDSHGAGYQMTSHFYALYNGGIWGRGLGNSITKKGYPETETDFIFSIITEE 282

30 Query: 312 LGLIGAGFILALVFFLILRIMNVGIKAKNPFNAMMALGVGGMMLMQVFNIGGISGLIPS 371
LGLICA +L L+F L +RI + + KN + LG G ++ +Q +N+G I+GL+P
Sbjct: 283 LGLIGALCVLFLFLSLCMRIFCLSSRCKNQAGLFLLGFGTLLFVQTIMNVGSIAGLMPM 342

30 Query: 372 TGVTFPFLSQGNSLLVLSVAVGFVLNIDASEKRDDI 408
TGV PF+S GG S L+LS+ +G LNI + + +++
Sbjct: 343 TGVPLPFVSYGGTSYLILSLGIGITLNISSKIQAEEI 379

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/166 (78%), Positives = 152/166 (91%), Gaps = 2/166 (1%)

35 Query: 1 MANSXYAMSGGWFGRLGNSIEKLGYLPEATTDFVFSIVIEELGVIGAGFILALVFFLI 60
+ANS YAMSGGWFG+GLGNSIEK GYLPEA TDFVFS+VIEELG+IGAGFILALVFFLI
Sbjct: 269 LANSYYAMSGGWFGQGLGNSIEKRGYLPEAQTDFVFSVVEELGLIGAGFILALVFFLI 328

40 Query: 61 LRIMHVGIAKADPFNSMTALGIGAMLLMQVFNIGGISGLIPSTGVTFPFLSQGNSLLV 120
LRIM+VGIAK+PFN+M+ALG+G M+LMQVFNIGGISGLIPSTGVTFPFLSQGNSLLV
Sbjct: 329 LRIMNVGIKAKNPFNAMMALGVGGMMLMQVFNIGGISGLIPSTGVTFPFLSQGNSLLV 388

45 Query: 121 LSAVIGFVLNIDANEKKELIMKAEERQYK--PQEKNEKIINLDAFK 164
LSVA+GFVLNIDA+EK++ I KEAE Y+ +++N K++N+ F+
Sbjct: 389 LSAVAVGFVLNIDASEKRDDIFKEAELSYRKDTRKENSKVVNKQFQ 434

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 319

A DNA sequence (GBSx0348) was identified in *Sagalactiae* <SEQ ID 1029> which encodes the amino acid sequence <SEQ ID 1030>. This protein is predicted to be probable cell division protein ftsW (ftsW). Analysis of this protein sequence reveals the following:

Possible site: 34

55 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.77	Transmembrane	12 - 28 (7 - 37)
INTEGRAL	Likelihood = -7.22	Transmembrane	76 - 92 (74 - 97)
INTEGRAL	Likelihood = -6.53	Transmembrane	182 - 198 (178 - 201)
INTEGRAL	Likelihood = -4.62	Transmembrane	51 - 67 (46 - 69)
60 INTEGRAL	Likelihood = -2.87	Transmembrane	202 - 218 (202 - 218)

-404-

----- Final Results -----

5 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9327> which encodes amino acid sequence <SEQ ID 9328> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA44490 GB:X62621 ORF2 N-terminal [Lactococcus lactis]
 Identities = 82/199 (41%), Positives = 122/199 (61%), Gaps = 9/199 (4%)

 Query: 1 MKIDKRHLNYSILIPYLILSILGLIVIIYSTTSATLIQLGANPFRSVINQGVFWAVSLVA 60
 M ++K + LNYSLIPYLIL+ +G+++I+STT +Q G NP++ VINQ F +S++
 15 Sbjct: 1 MNLNKNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGLNPYKLVINQTAFLVLSIIM 60

 Query: 61 IIFIYKLLNFKLNSKVLTMVAVLVEVFLLLIARF-----FTQEVNGAHGWIVIGPI-SF 113
 I IY+LKL LKN K++ + +++ + L+ R T VNGA GWI I I +
 20 Sbjct: 61 TAVIYRLKLRALKNRKMIGIIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTV 120

 Query: 114 QPAEYLKVIIVWYLAFTFARRQKKIEIYDQALTGKRWLPRSLSDLKDWRFYSLFMIGLV 173
 QPAE+ KV I+WYLA F+ +Q++IE D + KG+ L + L WR + ++ +
 25 Sbjct: 121 QPAEFAKVFIWYLASVFSTKQEEIEKNDINEIFKGKTLTQKL--FGWRLPVVAILLVD 178

 Query: 174 IAQPDGLNGSIIVLTVIIM 192
 + PDLGN II +IM
 30 Sbjct: 179 LIMPDLGNTMIIGAVLIM 197

There is also homology to SEQ ID 1028.

30 A related GBS gene <SEQ ID 8545> and protein <SEQ ID 8546> were also identified. Analysis of this protein sequence reveals the following:

 Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 15.18
 GvH: Signal Score (-7.5): -3.58
 Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 5 value: -9.77 threshold: 0.0
 INTEGRAL Likelihood = -9.77 Transmembrane 12 - 28 (7 - 37)
 INTEGRAL Likelihood = -7.22 Transmembrane 76 - 92 (74 - 97)
 40 INTEGRAL Likelihood = -6.69 Transmembrane 210 - 226 (201 - 227)
 INTEGRAL Likelihood = -6.53 Transmembrane 182 - 198 (178 - 201)
 INTEGRAL Likelihood = -4.62 Transmembrane 51 - 67 (46 - 69)
 PERIPHERAL Likelihood = 1.32 116
 modified ALOM score: 2.45

 *** Reasoning Step: 3

 ----- Final Results -----
 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 ORF02700(301 - 876 of 1377)
 EGAD|8615|8419(1 - 197 of 198) hypothetical protein in rpmg 3'region , fragment
 {Lactococcus lactis} SP|P27174|YRG2 LACLA HYPOTHETICAL PROTEIN IN RPMG 3'REGION (ORF2)
 (FRAGMENT). GP|44069|emb|CAA44490.1||X62621 ORF2 N-terminal {Lactococcus lactis}
 PIR|PC1134|PC1134 hypothetical protein 198 (rpmg 3' region) - Lactococcus lactis (fragment)
 %Match = 15.1
 60 %Identity = 42.3 %Similarity = 64.9
 Matches = 82 Mismatches = 64 Conservative Sub.s = 44

5

87 117 147 177 207 237 267 297
KA*I*Y*I**L*LVLILFLLPFFINFL*IYLTLGLND*NVPNSISN*SFI FVISIVGGYXX*LIXXXIMHNGNFLKY*RK*Y

327 357 387 417 447 477 507 537
NMKIDKRHLNLNYSILIPVYLILSLIGLIVIYSTTSATLIQLGANPFERSVINQGVFWAVLSVAIIIFTYKLKLNFLNKS KVL T
| : : | : ||||| ||||| : : | : : | : || : | | : : ||| | : : : | | : : || || | : :
MNLNKNNFLNYSILIPVYLILAGIGIVMIFSTTVPDQLQGLNPYKLVINOQTAFVLLSIIMI AVI YRLKLRALKNRKMIG

10

10 20 30 40 50 60 70

567 585 609 636 666 696 726 756
MAVLVEVFLLLLIARF----FT--QE VNGAHGWIVIGPI-SFPQAEYLKVII VWYLAFTFARRQKKIEIYDYQALTKGRWL
: : : : : | : | |||| ||| | | : ||| : || : |||| | : : : || | : || : |
IIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTVPQAEFAKFVI I WYLASVFSTKQEEIEKNDINEIFKGKTL

15

90 100 110 120 130 140 150

786 816 846 876 906 936 966 996
PRSLSDLKDWRFYSLFMIGLVLIAQPDLGNSSIIVLTVII MYCISGIGYRWFSALLGLIVVGSTLFIGTIAVVG VETMAKV
: | : | : : : : |||| || : ||
TQKL--FGGWRLPVVAILLVDLIMPDLGNMTMIIGAVALIMI

20

170 180 190

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 320

25 A DNA sequence (GBSx0349) was identified in *S.agalactiae* <SEQ ID 1031> which encodes the amino acid sequence <SEQ ID 1032>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

```

30      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3665 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1033> which encodes the amino acid sequence <SEQ ID 1034>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
```

```

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2373 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 35/41 (85%), Positives = 37/41 (89%)

50 Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACCYSPFY 41
 +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC S Y
 Sbjct: 1 LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQSRVY 41

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-406-

Example 321

A DNA sequence (GBSx0351) was identified in *S.agalactiae* <SEQ ID 1037> which encodes the amino acid sequence <SEQ ID 1038>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.65    Transmembrane    78 - 94 ( 78 - 95)
    INTEGRAL    Likelihood = -1.33    Transmembrane    421 - 437 ( 420 - 437)

----- Final Results -----
10  bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15  >GP:CAA00827 GB:A09073 phosphoenol pyruvate carboxylase
    [Corynebacterium glutamicum]
    Identities = 335/958 (34%), Positives = 539/958 (55%), Gaps = 80/958 (8%)

20  Query: 22  EIITEEVGLLKQLLDEATQKLIGSESFDKIE--KIVSLSLTD--DYTGLKETISALSNE 76
    + + +++ L Q+L E + G E ++ +E ++ S + + L + ++
    Sbjct: 3   DFLRDDIRFLGQILGEVIAEQEGQEVYELVEQARLTSFDIAGNAEMDSLQVFDGITPA 62

    Query: 77  EMVIVSRYFSLPLINISEDVDLAYEINYKNLNQDYLGLKLTST----TIDVV----- 125
    + ++R FS LL N++ED+ Y L + L T T+D
    Sbjct: 63  KATPIARAFSHFALLANLAEDL-----YDEELREQALDAGDTPPDSTILDATWLKLNKG 115

    Query: 126 -AGHENAKDILEHVNVPVLTHTPTQVQRKIVLELTSKIHDLLRKYRDVKAGIVNQ---- 180
    G E D+L + V PVLTAHPT+ +R+TV + I +R+ +++
    Sbjct: 116 NVGAEAADVLRNAEVAPVLTAHPTETRRRTVFDAQKWITTHMRERHALQSAEPTARTQS 175

30  Query: 181 --EKWYADLRRYIGIIMQTDITREKKLKVKNEITNVMEYINRSLIKAVTKLTAEYKALAA 238
    ++ ++RR I L+ QT IR + +++EI + YY SL++ + ++ +
    Sbjct: 176 KLDEIEKNIRRRITILNQ TALIRVARPRIEDEIEVGLRYYKLSLLEEIPRINRDVAVELR 235

    Query: 239 KK---GIHLENPKPLTM-GMWIGGDRDGNPFVTAETLRLSAMVQSEVIINHYIEQLNELY 294
    ++ G+ L KP+ G WIGGD DGNP+VTAET+ S +E ++ +Y QL+ L
    Sbjct: 236 ERFGEQVPL---KPVVKPGSWIGGDHGNPYVTAETVEYSTHRAAETVLKYYARQLHSLE 292

    Query: 295 RNMSLINLTVSPVLTLANQSQDNGSVYRENEPYRKAFNFIQDKLVQTLNLKVGSSPK 354
    +SLS + +V+P+L+ LA+ ++ R +EPYR+A + ++ +++ T
    Sbjct: 293 HELSLSDRMNKVTPQLLALADAGHNDVPSRVDEPYRRVHGVGRILAT----- 341

    Query: 355 EKFSVRQESSDIVGRIKSHIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYQGDS 414
    +++++G + + YA+ EEF D L + SL +
    Sbjct: 342 -----TAEELIGE-----DAVEGVWFKVFTPYASPEEFINDALTIDHSLRESKDVL 386

    Query: 415 LVDGELACLIQAVDIFGFYLATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQL 474
    + D L+ LI A++ FGF L +D+RQ+S E + EL + A + +Y LSE EK ++
    Sbjct: 387 IADDRLSVLISAIESFGFNLYALDLRQNSSEYEDVLTELFERAQVTANYRELSEAEKLEV 446

50  Query: 475 LLKELTEDPRTLSTTHAPKSELLQKELAIQTARELKDQLGEDIINQHIISHTESVSDMF 534
    LLKEL + SE+ +EL IF+TA E + G ++ IIS SV+D+
    Sbjct: 447 LLKELRSPRPLIPHGSDSEYSEVTDRELGIPTASEAVKFGPRMVPHCIISSMASSVTDVL 506

    Query: 535 ELAIMLKEVGLIDAN----QARIQIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNNN 590
    E ++LKE GLI AN + + ++PLFETIEDL I+ + +L + ++ +N
    Sbjct: 507 EPMVLLKEFGLIAANGDNPRGTVDVIPLFETIEDLQAGAGILDELWKIDLYRNYLLQRDN 566

    Query: 591 YQEIMLGYSDSNKDGGLSSGWTLYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGPSY 650
    QE+MLGYSDSNKDGGLY S+ W LY A+ +L ++ G+K+ FHGRGGTVGRGGGPSY
    Sbjct: 567 YQEVMLGYSDSNKDGGLYFSANWALYDAELQLVELCRSAGVKLRLFHGRGGTVGRGGGPSY 626

    Query: 651 EAITSQPFSGIKDRIRLTEQGEIENKYGNDAAYYNLEMLISASIDRMVTRMITNPNEI 710
    +AI +QP G+++ +R+TEQGEII KYGN + A NLE L+SA+++ + + +E+
  
```


-407-

Sbjct: 627 DAILAQPRGAVQGSVRITEQGEIISAKYGNPETAARNLEALVSATLE---ASLLDVSEL 682

Query: 711 DNFREITMDGIVSESNAV---YRNLVFDNPFYDYFFEASPIKEVSSLNIGSRPAARKTI 766
 + + D I+SE + + Y +LV ++ F DYF +++P++E+ SLNIGSRP++RK

5 Sbjct: 683 TDHQRAYD-IMSEISELSLKKYASLVHEDQGFIDYFTQSTPLQEIGSLNIGSRPSSRKQT 741

Query: 767 TEISGLRAIPWVFSWSQNRIMFPGWYGVGSFAFKHFI---EQDEANLAKLQTMQKWPFEN 823
 + + LRAIPWV SWSQ+R+M PGW+GVG+A + +I EQ +A+LQT+ + WFFF

10 Sbjct: 742 SSVEDLRAIPWVLSWSQSRVMLPGWFGVGTALQWIGEGEQATQRIAEQLTLNESWPFPT 801

Query: 824 SLLSNVDMVLSKSNMNIALQYAQLAGSKEVRD-VFNIILNEWQLTKDMILAIEQHDNLE 882
 S+L N+ V+SK+ + +A YA L EV + V++I E+ LTK M I D+LL+

15 Sbjct: 802 SVLDNMAQVMSKAELRLAKLYADLIPDTEVAERVYSVIREEYFLTKMFCVITGSDDLLD 861

Query: 883 ENPMLHASLDYRLPYFNVNLNYQIELIKRLSNQLEDDEYEKLIHITINGIATGLRNSG 940
 +NP+L S+ R PY LN +Q+E++R R E + I +T+NG++T LRNSG

Sbjct: 862 DNPLLARSVQRRYPYLLPLNVIQVEMMRRYRKGDQSEQVSRNIQLTMNGLSTALRNSG 919

20 A related GBS nucleic acid sequence <SEQ ID 10961> which encodes amino acid sequence <SEQ ID 10962> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1039> which encodes the amino acid sequence <SEQ ID 1040>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1613(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

30

An alignment of the GAS and GBS proteins is shown below:

Identities = 659/927 (71%), Positives = 779/927 (83%), Gaps = 11/927 (1%)

35 Query: 14 KLESSSNKEIITEEVGLLKQLLDEATQKLIGSESEFDKIEKIVLSLTDYDYLKETISAL 73
 KLESS+N++II EEV LLK++L+ T+++IG ++F IE I+ LS DY L++ ++ +

Sbjct: 5 KLESSNNQDIIEEVALLEKMLENITRRMIGDDAFIVIESIMVLSEKQDYIELEKVVANI 64

Query: 74 SNEEMVIVSRYFSILPLLINISEDVDLAYEINYKNNLNQDYLGLKLTSTIDVVAGHENAKD 133
 SN+EM ++SRYFSILPLLINISEDVDLAYEINY+NN + DYLGLK+ TI +AG +N KD

40 Sbjct: 65 SNQEMEVISRYFSILPLLINISEDVDLAYEINYQNNTDYLGLKALTIKDLAGKDNKGD 124

Query: 134 ILEHVNVPVLTAAHTQVQRKTIVLELTSKIHDLLRKYRDVKGAGVNVQEKWYADLRRYIGI 193
 ILE VNVVPVLTAAHTQVQRKT+LELT+ IH LLRKYRD KAG++N EKW +L RYI +

45 Sbjct: 125 ILEQVNVVPVLTAAHTQVQRKTILELTTHIHKLLRKYRDAKAGVINLEKWRQELRYIEM 184

Query: 194 IMQTDITIREKKLVKNEITNMEYNNRSLIKAVTKLTAEYKALAAKKGIHLENPKPLTMG 253
 IMQTD IREKKL+VKNEI NVM+YY+ SLI+AVTKLT EYK LA K G+ L+NPKP+TMG

Sbjct: 185 IMQTDIIREKKLQVKNKINVMQYDGLIQAVTKLTTEYKNLAQKHGLELDNPKPITMG 244

50 Query: 254 MWIGGDRDGNPFVTAETLRLSAMVQSEVIINHIEQLNELYRNMSLSINLTVSPPELVTL 313
 MWIGGDRDGNPFVTAETL LSA VQSEVI+N+YI++L LYR SLS L + + E+ L

Sbjct: 245 MWIGGDRDGNPFVTAETLCLSATVQSEVILNYYIDELAALYRTFSLSSTLVQPNSEVERL 304

55 Query: 314 ANQSQDNSVYRENEPYRKAFNFIQDKLVQTLNLKVGSSPKKFKVSRQESSDIVGRYIKS 373
 A+ SQD S+YR NEPYR+AF++IQ +L QT + L + + SS + S

Sbjct: 305 ASLQSDQSIYRGNEPYRRAFHYIQSRKQQTQIQLT-----NQPAASMSSSVGLNTSAWS 358

Query: 374 HIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYQDLSLVDGELACLIQAVDIFGFY 433
 A + + I AY + +FK DL ++QSL+ G +L++G+L ++QAVDIFGF+

60 Sbjct: 359 SPASLENPIL-----AYDSPVDFKADLKAIEQSLLDNGNSALIEGDLREVMAVDIFGFF 413

Query: 434 LATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQLLLKELTEDPRTLSSSTHAPK 493
 LA+IDMRQDSS+ EACVAELLK ANIVDDYSSLSE EKC +LL++L E+PRTLSS K

-408-

Sbjct: 414 LASIDMRQDSSVQEACVAELLKGANIVDDYSSLSETEKCDVLLQQLMEEPRTLSSAAVAK 473
 Query: 494 SELLQKELAI FQTARELKDQLGEDI INQHIISHTESVSDMFELAIMLKEVGLIDANQARI 553
 S+LL+KELAI+ TARELKD+LGE++I QHIISHTESVSDMFELAIMLKEVGL+D +AR+
 5 Sbjct: 474 SLLLEKELAIYTTARELKDKLGEVVKQHIISHTESVSDMFELAIMLKEVGLVDQQRARV 533
 Query: 554 QIVPLFETIEDLDNSRDIMTQYLHYELVKKIATNNNYQEIMLGYSNSKDGGLSSGWT 613
 QIVPLFETIEDLDN+RDI YL +++VK WIATN NYQEIMLGYSNSKDGGL+SGWT
 10 Sbjct: 534 QIVPLFETIEDLDNARDIMAAYLSDIVKSWIATNRNYQEIMLGYSNSKDGGLASGWT 593
 Query: 614 LYKAQNELTKIGEENGKITFFHGRGGTVGRGGGPSYEAITSQPFSGIKDRIRLTEQGEI 673
 LYKAQNELT IGEE+G+KITFFHGRGGTVGRGGGPSY+AITSQPFSGIKDRIRLTEQGEI
 Sbjct: 594 LYKAQNELTAIGEENGKITFFHGRGGTVGRGGGPSYDAITSQPFSGIKDRIRLTEQGEI 653
 15 Query: 674 IENKYGNDAAAYYNLEMLISASIDRMVTRMITNPNEIDNPRETMDGIVSESNAVYRNLFV 733
 IENKYGND AYY+LEMLISASI+RMVT+MIT+PNEID+FRE MD IV++SN +YR LVF
 Sbjct: 654 IENKYGNDVAYYHLEMLISASINRMVTQMITDPNEIDSPREIMDSIVADSNIYRKLVF 713
 Query: 734 DNPYFYDYFFEASPIKEVSSLNIGSRPAARKTITEISGLRAIPWVFSWSQNRIMFPGWYG 793
 DNP+FYDYFFEASPIKEVSSLNIGSRPAARKTITEI+GLRAIPWVFSWSQNRIMFPGWYG
 20 Sbjct: 714 DNPYFYDYFFEASPIKEVSSLNIGSRPAARKTITEITGLRAIPWVFSWSQNRIMFPGWYG 773
 Query: 794 VGSFAKHFIEQDEANLAKLTMYQKWPFFNSLLSNVDMVLSKSNMNIALQYAQLAGSKEV 853
 VGSFAK +I++ + NL +LQ MYQ WFFF+SLLSNVDMVLSKSNMNIA QYAQLA ++V
 25 Sbjct: 774 VGSFAKRYIDRAQGNLERLQHMVQWFFFHSLLSNVDMVLSKSNMNIAFYAQLAERQDV 833
 Query: 854 RDVFNIILNEWQLTKDMILAEQHDNLEENPMLHASLDYRLPYFNVNLNYQIELIKRLR 913
 RDVF IL+EWQLTK++ILAI+ HD+LLE+NP L SL RLPYFNVNLNY+QIELIKR R
 30 Sbjct: 834 RDVFYILDEWQLTKNVILAIQHDHDLLEDNPSLKHSLKSLRPLPYFNVNLNYQIELIKRWR 893
 Query: 914 SNQLDEDEYKLIHITTINGIATGLRNSG 940
 +NQLDE+ EKLIH TINGIATGLRNSG
 Sbjct: 894 NNQLDENDEKLIHTTINGIATGLRNSG 920

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 322

A DNA sequence (GBSx0352) was identified in *S.galactiae* <SEQ ID 1041> which encodes the amino acid sequence <SEQ ID 1042>. This protein is predicted to be *Bacillus licheniformis* Pz-peptidase
 40 homologue (pepF). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.3012(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1043> which encodes the amino acid
 50 sequence <SEQ ID 1044>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3137(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-409-

Identities = 512/593 (86%), Positives = 564/593 (94%)

Query: 1 MKLKKRSEFFPENELWDLTALYKDRQDFLLAIEKALEDIKVFKNYEGKLNCEVEDFTSALM 60
 M+LKKRSEFFPENELWDLTALYKDRQDFLLAIEKAL+DI +FK+NYEG+L V+DFT AL+
 5 Sbjct: 26 MELKKRSEFFPENELWDLTALYKDRQDFLLAIEKALQDIDLFKRNYEGRLTSVDDFTQALI 85

Query: 61 EIEHIYIQMSHIDTYAFMPQTTDFSNEEFAQISQAGSDFATKANVLLSFFNTALANADIK 120
 EIEHIYIQMSHI TYAFMPQTTDFS+E FAQI+QAG DF TKA+V LSFF+TALANAD+
 10 Sbjct: 86 EIEHIYIQMSHIGTYAFMPQTTDFSDESFAQIAQAGDDFMKASVALSFFDTALANADLD 145

Query: 121 ILDSLNNPHFKATIRQAKIQKHLLSPEVEKALTNLNEVLNTPYDIYTKMRAGDFDMED 180
 +LD+LE NP+F A IR AKIQK+HLLSP+VEKAL NL EV+N PYDIYTKMRAGDFDM+D
 Sbjct: 146 VLDLTLEKNPYFSSAIRMAKIQKEHLLSPDVEKALANLREVINAPYDIYTKMRAGDFDMDD 205

Query: 181 FEVDGKTYKNSFVYENYFQNHENAEIREKSFPSFKGLRKHQNAAYLAKVKSEKLI 240
 FEVDGKTYKNSFV+YEN++QNHENAEIREK+FRSFSKGLRKHQN AAAYLAKVKSEKLI+
 15 Sbjct: 206 FEVDGKTYKNSFVSYENFYQNHENAEIREKAFRSFSKGLRKHQNTAAAYLAKVKSEKLI 265

Query: 241 ADMRGYDSVFDYLLSEQEVDRLSMFDRQIDILIMDEFQGPVQAFKFLKHIADVNGIEKMTFADW 300
 ADM+GY SVFDYLL+EQEVDRLS+MDFDRQIDILIM EFGPVAQ+FLKH+A VNG+EKMTFADW
 20 Sbjct: 266 ADMKGYASVFDYLLAEQEVDRSLFDRQIDILIMTEFGPVAQFLKHVAQVNGLEKMTFADW 325

Query: 301 KLDIDNENLNEVSIINDAYDLVMKSVAPLGKEYSQEVERYQKERWVDFANANKDSGGYAA 360
 KLDIDN+LNPEVSI+ AYDLVMKS+APLG+EY++E+ERYQ ERWVDFANANKDSGGYAA
 25 Sbjct: 326 KLDIDNENLNEVSIIDGAYDLVMKSLAPLGQEYTKIERYQTERWVDFANANKDSGGYAA 385

Query: 361 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSFFNTHMSTYYVEAPSTFNE 420
 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQS+FNTNTHMSTYYVEAPSTFNE
 30 Sbjct: 386 DPYKVHPYVLMWSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSYFNTNTHMSTYYVEAPSTFNE 445

Query: 421 LLLSDYLENQFDTARQKRFALAHRLTDTYFHNFTIHLLEAAAFQKRVYTLIEEGGTFGAEQ 480
 L+LSDYLE+QFD RQKRFALAHRLTDTYFHNFTIHLLEAAAFQKRVYTLIEEGGTFGA+Q
 Sbjct: 446 LMLSDYLEHQFDDPRQKRFALAHRLTDTYFHNFTIHLLEAAAFQKRVYTLIEEGGTFGADQ 505

Query: 481 LNAIMKEVLTDQFWGDAIEIDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNKNNP 540
 LNA+MKEVLTDQFWGDA++IDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNK+NP
 35 Sbjct: 506 LNAMMKEVLTDQFWGDAVIDDDAALTWMRQAHYMGLYSYTYSAGLVISTAGYLNKNNP 565

Query: 541 NGAKEWLAFLKSGGSRTPLLETALLISADISTDKPLRDTINFLSNTVDQIINYS 593
 NGAKEWL FLKSGGSRTPL+TA+LI ADI+T+KPLRDTI FLS+TVDQII+Y+
 40 Sbjct: 566 NGAKEWLDLFLKSGGSRTPLDTAMLIGADIATEKPLRDTIQFLSNTVDQIISYT 618

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 323

A DNA sequence (GBSx0353) was identified in *S. galactiae* <SEQ ID 1045> which encodes the amino acid sequence <SEQ ID 1046>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1047> which encodes the amino acid sequence <SEQ ID 1048>. Analysis of this protein sequence reveals the following:

Possible site: 19

-410-

>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 72/127 (56%), Positives = 85/127 (66%)

Query: 1 MKKYIKLFLLTVPATTLVACGQPSTSNKTTTSSSTLEVKGVELVVKEDTNVLSEKVVYHKG 60
 + K K L + A LVAC Q + +TT S V LVVKEDTN + EKV + KG
 Sbjct: 1 VNKRFKTGFLALVAMLLVACSGQGTKQIQITPSVFKADHHVRLVVKEDTNTVDEKVSFGKG 60

15 Query: 61 DTVLDVLKANYKVKEKDGFIITSIDGISQDETGLYWMFKVNNKLAPKAANQIKVKKNDKI 120
 DTVL+VLK NY+VKEKDGFIIT+IDGI QD YW+FKVN K+A K A+QI VK D I
 Sbjct: 61 DTVLEVLKDNYEKEDGFIITAIDGIEQDTKANKYWLKFNKMGADKGADQITVKDGD SI 120

20 Query: 121 EFYQEVY 127
 EFYQEV+
 Sbjct: 121 EFYQEVF 127

25 SEQ ID 1046 (GBS185) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 6; MW 15.7kDa).

GBS185-His was purified as shown in Figure 199, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 324

30 A DNA sequence (GBSx0354) was identified in *S.agalactiae* <SEQ ID 1049> which encodes the amino acid sequence <SEQ ID 1050>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

35 INTEGRAL Likelihood = -4.46 Transmembrane 75 - 91 (67 - 94)
 INTEGRAL Likelihood = -4.41 Transmembrane 33 - 49 (30 - 49)
 INTEGRAL Likelihood = -2.60 Transmembrane 53 - 69 (52 - 70)
 INTEGRAL Likelihood = -1.38 Transmembrane 108 - 124 (106 - 124)
 INTEGRAL Likelihood = -0.06 Transmembrane 149 - 165 (149 - 165)

40 ----- Final Results -----

 bacterial membrane --- Certainty=0.2784 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9731> which encodes amino acid sequence <SEQ ID 9732> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10929> which encodes amino acid sequence <SEQ ID 10930> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1051> which encodes the amino acid sequence <SEQ ID 1052>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have a cleavable N-term signal seq.

-411-

```

INTEGRAL    Likelihood = -7.96    Transmembrane    50 - 66 ( 49 - 71)
INTEGRAL    Likelihood = -5.73    Transmembrane    101 - 117 ( 99 - 124)
INTEGRAL    Likelihood = -4.41    Transmembrane    141 - 157 ( 139 - 159)
INTEGRAL    Likelihood = -4.25    Transmembrane    73 - 89 ( 67 - 92)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/163 (50%), Positives = 120/163 (73%), Gaps = 3/163 (1%)

```

Query: 10  LTRVAILSALCVVLRVAFAPLPNIQPTAIFLITVVLFDLKEGVATVTITMLVSSFLMGF 69
          ++R+AI+SALCVVLR  F+ LPN+QP+TA  L  ++ F L E V  + + + +S+FL+GF
Sbjct: 6   MSRIAISALCVVLRMVFSSLPNVQPVTAFLLSYLLYFGLAEAVLVMMLCLFLSAFLLG 65

Query: 70  GPWVFLQIISFTLILCLWKFLIYPLTKAVCFGKITEVVLTQFFAGGLGVVYGVIIDTCFA 129
          GPWVF Q+  F L+L LW+F++YPL++  F K ++  Q F  G++YGV+IDTCFA
Sbjct: 66  GPWVFWQVTCFVLVLLWRFVLYPLSQQ--FPKY-QLGCQAFVLVALCGLLYGVLIIDTCFA 122

Query: 130 WLYHMPWWTYVVLAGLSFNMAHALSTCLFYPLLLPILRRFRNEK 172
          +LY MPWW+YVLAG+ FN+AHALST +F+P+++ + RR  E+
Sbjct: 123 YLYSMFWWSYVLAGMPFNIAHALSTLVFPVVMMLFRRLIGE 165

```

A related GBS gene <SEQ ID 8549> and protein <SEQ ID 8550> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 10
McG: Discrim Score: 6.79
GvH: Signal Score (-7.5): -0.91
Possible site: 28

```

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 3 value: -4.46 threshold: 0.0

```

INTEGRAL    Likelihood = -4.46    Transmembrane    35 - 51 ( 29 - 54)
INTEGRAL    Likelihood = -1.38    Transmembrane    68 - 84 ( 66 - 84)
INTEGRAL    Likelihood = -0.06    Transmembrane    109 - 125 ( 109 - 125)
PERIPHERAL  Likelihood = 7.53      88
modified ALOM score: 1.39

```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.2784(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01220(421 - 552 of 1002)
GP|9950155|gb|AAG07353.1|AE004814_8|AE004814(16 - 56 of 69) hypothetical protein
{Pseudomonas aeruginosa}
%Match = 3.2
%Identity = 39.5 %Similarity = 60.5
Matches = 17 Mismatches = 15 Conservative Sub.s = 9

222      252      282      312      342      372      402      432
STLTKLTRVAILSALCVVLRVAFAPLPNIQPTAIFLITVVLFDLKEGVATVTITMLVSSFLMGFGPWVFLQIISFTLIL
                                                                |:::
                                                                MDPELFEEWMMTGLVTVLI
                                                                10

462      492      522      552      582      612      642      672
CLWKFLIYPLTKAVCFGKITEVVLTQFFAGGLGVVYGVIIDTCFAWLYHMPWWTYVVLAGLSFNMAHALSTCLFYPLLLPT

```

-412-

```

      : |:: | |   ||   ::| ||| ||| | | |
LFMAFIVWDLAKSKAGKFGTLIL--FFALGLGV-LGFIKGLVIGSLEGAGM
      30      40      50      60

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 325

- A DNA sequence (GBSx0355) was identified in *S.agalactiae* <SEQ ID 1053> which encodes the amino acid sequence <SEQ ID 1054>. This protein is predicted to be endolysin. Analysis of this protein sequence
10 reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

```

- ```

----- Final Results -----
15 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT<sup>1</sup> database:

- ```

20 >GP:CAA72266 GB:Y11477 endolysin [Bacteriophage Bastille]
      Identities = 64/210 (30%), Positives = 95/210 (44%), Gaps = 15/210 (7%)

Query: 66 KPIIDVSGWQLPKEIDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKTHIKEF 125
      K I+D+S      +ID+DT      +S + R  G + + +N      +D+ +KT +
25 Sbjct: 12 KTIVDISHHNA--DIDFDTAKNYVSMFIARTGDGHRYN--SNGELQGVVDRKYKTFVANM 67

Query: 126 QKRNPVAVYSYALGSSVKEMKEEAQIFYKNAAPYKPTFFYWIDVEETMSNMNKGVQAFR 185
      + R IP  Y +   S V  K+EA+ F+ N      T + D E T  NM + +Q F
30 Sbjct: 68 KARGIPFGNYMFNRFSGVASAKQEAEFFW-NYGDKDATVWVCDAEVSTAPNMKECIQVFI 126

Query: 186 KELKRLGAKNVGIYIGTYFMTEQGISVKGFDVWIPYTGSDSGYYEAPQTELKYDLHQY 245
      LK LGAK VG+YIG +   E G      D  WIP YG+   +      DL Q+
35 Sbjct: 127 DRLEKELGAKKVGLYIGHHKYQEFGGKDVNCDFTWIPRYGNKPAF-----ACDLWQW 177

Query: 246 TSQGYLPGFNQPLDLNQIAVNKDKKKTYEK 275
      T  G + G + D+N + +K      EK
40 Sbjct: 178 TEYGNIAIGIK-CDINVLYGDKPMSFFTEK 206

```

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1055> which encodes the amino acid
40 sequence <SEQ ID 1056>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -16.98      Transmembrane      8 - 24 ( 3 - 28)

```

- ```

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.7793(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

- 50 An alignment of the GAS and GBS proteins is shown below:

```

 Identities = 198/278 (71%), Positives = 235/278 (84%)

```

- ```

Query: 1  MRRRIKPIVVAVFFSLFGLLLIIGHLSTNTLKKELVEAKKTIPSVKASKVPQKSTSSKD 60
      MRR+IKPIVV VFF L  ++LIIG  + + +KE+ +AK IP  ++      K+++S+
55 Sbjct: 1  MRRKIKPIVVLVFFILLAMVLIIGKRQANHAKQKEVEDAKSHIPIATSNPGKAKTSTSET 60

Query: 61  KEFVLKPIIDVSGWQLPKEIDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKT 120
      ++F+L PI+DVSGWQLP+EIDYDTLS++ISG ++RV+GGS+I+  NNAA+TTGIDKSFKT

```

10

20

25

30

The protein has homology with the following sequences in the databases:

35

Bacteriophage

40

45

50

55

60

65

PAFACDLWQWTEYGNIAIGIK-CDINVLYGDKPMSFFTEKEGAKETLVPALNKVVITYEVGTNLIPEIQDKLAFLGYEARL
180 190 200 210 220 230 240

SEQ ID 8552 (GBS206) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
5 extract is shown in Figure 51 (lane 6; MW 31.7kDa).

GBS206-His was purified as shown in Figure 206, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 326

10 A DNA sequence (GBSx0356) was identified in *S.galactiae* <SEQ ID 1057> which encodes the amino
acid sequence <SEQ ID 1058>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence
15 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (183 - 200)
----- Final Results -----
bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9729> which encodes amino acid sequence <SEQ ID 9730>
was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:AAG20117 GB:AE005090 NADH dehydrogenase/oxidoreductase-like
protein; Nola [Halobacterium sp. NRC-1]
Identities = 38/156 (24%), Positives = 83/156 (52%), Gaps = 13/156 (8%)
Query: 19 TMEILIAAGSGFLGKQIIKAALTGKHVAYLSRHEGKGDIFKDPRLTYIRGDITEADKIH 78
+M++L+ GG+GF+G + + +GH V +R + D +T I GD+T + +
30 Sbjct: 8 SMDVLNTGGTGFTGTHLCRELDGRHDVTAFAREPADAALPAD--VTRIVGDVTVKETVA 65
Query: 79 LEDRTFDILIDCIGA---IKPNQLD----ELNVKATQKAVLCHKNQIPKLVYISA----- 127
D +++ + KP+ D ++++ T+ VA + + ++ +SA
35 Sbjct: 66 NAIDGHDAVNLVALSPLFKPSGGDSRHLVDVHLGGTENVVAAASEAGVEYILQLSALDAD 125
Query: 128 NSGYSAYIKSKRKAEQIIKASGLDYLFVRPGLMYGE 163
+G +AY+++K +AE+ +++S L + VRP +++G+
Sbjct: 126 PTGPTAYLRAKGRAEEAVRSSDLHHTIVRPSVVFGE 161

40 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8553> and protein <SEQ ID 8554> were also identified. Analysis of this
protein sequence reveals the following:

Lipop Possible site: -1 Crend: 5
McG: Discrim Score: -7.99
45 GvH: Signal Score (-7.5): -6.34
Possible site: 41
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -1.44 threshold: 0.0
INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (183 - 200)
50 PERIPHERAL Likelihood = 4.29 20
modified ALOM score: 0.79

*** Reasoning Step: 3

Example 327

A DNA sequence (GBSx0357) was identified in *S.galactiae* <SEQ ID 1059> which encodes the amino acid sequence <SEQ ID 1060>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2850(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC36853 GB:L23802 regulatory protein [Enterococcus faecalis]
 Identities = 61/164 (37%), Positives = 96/164 (58%), Gaps = 13/164 (7%)

Query: 1 MSKKNKIKKTLVDQILDKAKIEH-----DSLQLDALQGDLPNGIQKDIFKTLALI 51
 M+KK +KT +++++ K+ + D L +++ L GI+K IFKTL +
 Sbjct: 1 MAKKKTQQKTNAMRMVEQHKVPYKEYEFAWSEDLHSAESVAESL--GIEKGRIFKTLVTV 58

Query: 52 GDKTGPIIGILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKH 111
 G+KTGP++ +PP + L KKLAK SGNKKV+M+ KDL+ TGYI G +P G+ K
 Sbjct: 59 GNKTGEVAVIPGNQELDLKKLAKASGNKKVEMHLKDLLEATTGYIRGGCSPTGM--KKQ 116

Query: 112 YPIFIDTIALEKQELIVSAGEIGRSIRINSEVLADFVNKAFADI 155
 +P ++ A + +IVSAG+ G I + E + N +FA+I
 Sbjct: 117 FPTYLAEEAQQYSALIVSAGKRGMQIELAFAELSLINGQFAEI 160

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1061> which encodes the amino acid sequence <SEQ ID 1062>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 114/157 (72%), Positives = 139/157 (87%)

Query: 1 MSKKNKIKKTLVDQILDKAKIEHDSLQLDALQGDLPNGIQKDIFKTLALIGDKTGPIIG 60
 M+KK K+KKTLLV+QILDKA I H L+L+AL+GD P+ +Q DI+KTLAL GD+TGP+IG
 Sbjct: 1 MAKKTKLKKTLVEQILDKANIAHQGLKLNALGDFPDDLQPSDIYKTLALTGDQTGPLIG 60

Query: 61 ILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKHNYPIFIDTIA 120
 I+PLTEHLSEK+LAK+SGNKKV M+PQKDLQK TGYIHGANNP+GIRQKH+YPIFID A
 Sbjct: 61 IIPLTEHLSEKQLAKVSGNKKVSMVPQKDLQKTTGYIHGANNPVGIRQKHSYPIFIDQTA 120

Query: 121 LEKQELIVSAGEIGRSIRINSEVLADFVNKAFADIKE 157
 LEK ++IVSAGE+GRSI+I+S+ LADFV A FAD+K+
 Sbjct: 121 LEKGQIIIVSAGEVGRSIKISSQALADFVGASFADLKK 157

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 328

A DNA sequence (GBSx0358) was identified in *S.galactiae* <SEQ ID 1063> which encodes the amino acid sequence <SEQ ID 1064>. Analysis of this protein sequence reveals the following:

-417-

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4719(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8555> which encodes amino acid sequence <SEQ ID 8556>
 10 was also identified. This protein belongs to the glycolysis/gluconeogenesis pathway, and such proteins have
 been experimentally detected as surface-exposed in *Streptococci*. The protein has homology with the
 following sequences in the GENPEPT database:

>GP:AAD36444 GB:AE001791 phosphoglycerate mutase [Thermotoga maritima]
 Identities = 65/191 (34%), Positives = 93/191 (48%), Gaps = 13/191 (6%)

15 Query: 5 MKFYLV RHGKTQWNLEGRFQ GANGDSP LLEEAEIEEELGQYLSSIHFDAVYSSDLGRAR 64
 MK YL+RHG+T WN +G +QG D PL E E+ +L L + DA+YSS L R+
 Sbjct: 1 MKLYLIRHGETIWNKGLWQGV T-DVPLNERGREQARKLANSLKRV--DAIYSSPLKRSL 57

20 Query: 65 DTVNILNDANSCPK EIHYPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124
 +T + A KEI LRE + G + YP + + +P M
 Sbjct: 58 ETAEEI--ARRFEKEIIVEEDLRECEISLWNLTVEEAIREY PVEFKWSSDP---NFGM 112

25 Query: 125 FGAESLYQTTHRVESFLRSLASK----NYDKVLIVGHGANLTASIRSLGQYQYGS LHYKD 180
 G ES+ +RV + +S+ + V+IV H +L A I +LG LH
 Sbjct: 113 EGLESMRNVQNRVVKAIMKIVSQEKLNGSENVVIVSHSLSLRAFICWILGLPL-YLHRNF 171

 Query: 181 KLDNASLTIIE 191
 KLDNASL+++E

30 Sbjct: 172 KLDNASLSVVE 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1065> which encodes the amino acid
 sequence <SEQ ID 1066>. Analysis of this protein sequence reveals the following:

Possible site: 24
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 127/205 (61%), Positives = 152/205 (73%)

45 Query: 5 MKFYLV RHGKTQWNLEGRFQ GANGDSP LLEEAEIEEELGQYLSSIHFDAVYSSDLGRAR 64
 MK Y VRHGKT WNLEGRFQGA GDSPLLEEAE +E+ LG+ L+ + FDAVY+SDL RA
 Sbjct: 1 MKLYFVRHGKTLWNLEGRFQAGGDSPLLEEAKDEIHLLGKELAKVAFDAVYTSDLQRAM 60

50 Query: 65 DTVNILNDANSCPK EIHYPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124
 T I+ DA ++++T QLREW LG LEG KIATM AIYP+QM AF +N QFK D
 Sbjct: 61 ATAAIIIDAFDQQPKLYHTDQLREWRLGKLEGAKIATMAAIYPQQLAFRENLAQFKPDQ 120

 Query: 125 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGQYQYGS LHYKDKLDN 184
 F AES+YQTT RV ++S K+Y V LIVGHGANLTA+IRSLG++ L K LDN
 55 Sbjct: 121 FEAESIYQTTQRVCHLIQSFKDKHYQNV LIVGHGANLTATIRSLGFEPA LLLAKGGLDN 180

 Query: 185 ASLTIETHDFKDFNCLTWNDKSYL 209
 ASLTI+ET D+ ++CL WNDKS+L
 60 Sbjct: 181 ASLTILETKDYLTVDCLIWNDKSFL 205

SEQ ID 8556 (GBS314) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 4; MW 27.2kDa), in Figure 169 (lane 15-17; MW 41.6kDa) and in Figure 239 (lane 4; MW 41.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 4; MW 52.1kDa).

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 329

A DNA sequence (GBSx0359) was identified in *S.agalactiae* <SEQ ID 1067> which encodes the amino acid sequence <SEQ ID 1068>. Analysis of this protein sequence reveals the following:

10 Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.3014(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:CAB12562 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
Identities = 69/232 (29%), Positives = 108/232 (45%), Gaps = 9/232 (3%)

Query: 4 SIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSIDIGFPRVMAGEWITEY 63
+++FDVDDTI D QA +A+ F D ++ N +++ + + G+ T +
25 Sbjet: 6 TLLFDVDDTILDFQAAEALALRLLFEDQNIPLTNDMKAQYKTINQGLWRAFEEGKMTRDE 65

Query: 64 FRFWRKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDLFLKSKNVPNGII 123
R L E+GY EA G ++ Y LE L + L + + I+
30 Sbjet: 66 VVNTFRSALLKEYGY---EADGALLEQKYRRFLKEGHQLIDGAFDLISNLQQQFDLYIV 121

Query: 124 TNGPTEHQKLVKKLGLYDVDPKRVIVSQATGFQKPEKEIFNLAAEQF-DMNPSTTLIV 182
TNG + Q K+++ GL+ + K + VS+ TGFQKP KE FN E+ + TL +
35 Sbjet: 122 TNGVSHTQYKRLRDSGLFPFF--KDIFVSEDTGFKPMKEYFNVVFERIPQFSAEHTLII 179

Query: 183 GDSYDNDIMGAFNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKV 234
GDS DI G G + W N + P I P Y+ I E+L+ + +
Sbjet: 180 GDSLTAIDIKGGQLAGLDTCCWNPDMKPNVPEIPTYE--IRKLEELYHILNI 229

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1069> which encodes the amino acid sequence <SEQ ID 1070>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.3216(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 276/300 (92%), Positives = 292/300 (97%)

50 Query: 1 MITSIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSIDIGFPRVMAGEWT 60
MIT+IVFDVDDTIYDQQAPYRIA+EKCFPDFDMS +NQAYIRFRHYSID+GFPRVMAGEWT
Sbjet: 1 MITAIVFDVDDTIYDQQAPYRIAMEKCFPDFDMSVMNQAYIRFRHYSIDVGFPRVMAGEWT 60

55 Query: 61 TEYFRFWRKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDLFLKSKNVP 120
TEYFRFWRKETLLEFGYREIDEA G++FQE+YEHELENITMLDEMRTLDLFLKSKNVP

-419-

Sbjct: 61 TEYFRFWRKETLLEFGYREIDEAAGVHFQEVYEHELENITMLDEMRTLDLFLKSKNVPM 120

Query: 121 GIITNGPTEHQLKKVKKLGLYDYDVKRVIVSQATGFQKPEKEIFNLAAEQFDMNPSTTL 180
GIITNGPTEHQLKKV+KLGLYDY+D KRVIVSQATGFQKPEKEIFNLAAEQFDMNP TTL

5 Sbjct: 121 GIITNGPTEHQLKKVRKLGLYDYIDAKRVIVSQATGFQKPEKEIFNLAAEQFDMNPQTTL 180

Query: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKVLFDLPD 240
YVGDSYDNDIMGAFNGGWHSMWFNHRGR LKPG KPVYDVAIDNFEQLFGAVKVLFDLPD

10 Sbjct: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRQLKPGTKPVYDVAIDNFEQLFGAVKVLFDLPD 240

Query: 241 NKFIFDINDKSNPVLEMGLNGLMMAAERLLESNMSVDKVVILLRLTAKQEKVLRMKYAR 300
NKFIFD+NDK NP+L+MG+NNGLMMAAERLLESNMS+DKVVILLRLT +QEKVLR+KYAR

Sbjct: 241 NKFIFDVNDKKNPILQMGINGLMMAAERLLESNMSIDKVVILLRLTKQEKVLRMLKYAR 300

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 330

A DNA sequence (GBSx0360) was identified in *S. agalactiae* <SEQ ID 1071> which encodes the amino acid sequence <SEQ ID 1072>. Analysis of this protein sequence reveals the following:

20 Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9727> which encodes amino acid sequence <SEQ ID 9728> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB11858 GB:Z99104 lysyl-tRNA synthetase [Bacillus subtilis]
Identities = 318/490 (64%), Positives = 390/490 (78%), Gaps = 1/490 (0%)

35 Query: 44 EELNDQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDLHDIEETATIA 103
EELNDQ VRR+KM L + GIDPFG RFERT S ++ Y D +KE+L + TIA

Sbjct: 9 EELNDQLQVRRDKMNQLRDNGIDPFGARFERTHQSQEIVISAYQDLTKEELEKAIEVTIA 68

Query: 104 GRLMTKRGKGVGFAGHIDREGQIQIYVRKDSVGEENYEIFKKADLGDFLGVEGQVMRTD 163
GR+MTKRGKGVGFAGH+QD EGQIQIYVRKDSVG++ YEIFK +DLGD +GV G+V +T+

40 Sbjct: 69 GRMMTKRGKGVGFAGHLDLEGGQIQIYVRKDSVGDDQYEIFKSSDLGLIGVTGKVFKTN 128

Query: 164 MGELSIKATHITHLSKALRPLPEKFHGLTDIETIYKRHLDLISNRDSFDRFVTRSKIIS 223
+GELS+KAT L+KALRPLP+K+HGL D+E YR+R+LDLI N DS F+TRSKII

45 Sbjct: 129 VGELSVKATSFELLTKALRPLPDKYHGLKDVEQRYRQRYLDLIVNPD SKHTFITRSKIIQ 188

Query: 224 EIRRFMDNSGFLVETPVLHNEAGGASARPFITHHNAQDIDMVLRIATELHLKRLIVGGM 283
+RR++D +G+LEVETP +H+ GGASARPFITHHNA DI + +RIA ELHLKRLIVGG+

Sbjct: 189 AMRRYLDHGYLEVETPTMHSIPGGASARPFITHHNAIDIPLYMRIAELHLKRLIVGGL 248

50 Query: 284 ERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVTKTIVKGDGPIN 343
E+VYEIGR+FRNEG+ HNPEFT IE Y+AYADY+DIM LTE ++ H+ + V G I

Sbjct: 249 EKVYEIGRVFRNEGVSTRHNPEFTMIELYEAYADYKDIMSLTENLVAHIAQEVLTGTTTIQ 308

55 Query: 344 YQGTEIKINEPFKRVHMVDAVKEITGIDFWKEMTLEEAQALAEKNVPLEKHFTTTVGHII 403
Y +I + +KR+HMVDAVKE TG+DFW+E+T+E+A+ A+E V + K TVGHII

Sbjct: 309 YGEEQIDLKPEWKRIHMVDAVKEATGVDFWEEVIVEQAREYAKEHEVEI-KDSMTVGHII 367

Query: 404 NAFFEFVEDTLIQPTFVFGHPVEVSPLAKKNDTPRFTDRFELFIMTKEYANAFTELND 463
N FFE+ +E+TLIQPTF++GHPVE+SPLAKKN DPRFTDRFELFI+ +E+ANAFTELND

60 Sbjct: 368 NEFFEQIEETLIQPTFIYGHVPVEISPLAKKNPEDPRFTDRFELFIVGREHANAFTELND 427

-420-

Query: 464 PIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGMPPPTGGLGIGIDRLCMLLTDTTTI 523
 PIDQ RFEAQ +E G+DEA +D D+VEALEYGMPPPTGGLGIGIDRL MLLT+ +I
 Sbjct: 428 PIDQRRERFEAQLKEREAGNDEAHLMDDEFVEALEYGMPPPTGGLGIGIDRLVMLLTNAPSI 487

Query: 524 RDVLLFFPTMK 533
 RDVLLFP M+
 Sbjct: 488 RDVLLFFQMR 497

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1073> which encodes the amino acid sequence <SEQ ID 1074>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4694 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below:

Identities = 439/500 (87%), Positives = 474/500 (94%)

Query: 34 LEEIMSNQHIEELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDI 93
 LEE MSNQHIEELNDQQIVRREKM AL EQGIDPFGKRF+RTA S +L EKYADK+KE+L
 Sbjct: 1 LEENMSNQHIEELNDQQIVRREKMTALAEQGIDPFGKRFDRRTANSABLKKEKYADKTKEEL 60

Query: 94 HDIETATIAGRIMTKRGKGVGFHAIQDREGQIQIYVRKDSVGEENYEIIFKKADLGDFL 153
 H++ ETA +AGRLMTKRGKGVGFH+QDREGQIQ+YVRKDSVGE+NYEIFFKADLGDF+
 Sbjct: 61 HELNETAIVAGRLMTKRGKGVGFHQLQDREGQIQLYVRKDSVGEDNYEIFFKADLGDFI 120

Query: 154 GVEGQVMRTDMGELSIKATHITHLSKALRPLPEKFHGLTDIETIYRKRHLDLISNRDSD 213
 GVEG+VMRTDMGELSIKAT +THLSK+LRPLPEKFHGLTDIETIYRKRHLDLISNR+SFD
 Sbjct: 121 GVEGEVMRTDMGELSIKATKLTHLSKSLRPLPEKFHGLTDIETIYRKRHLDLISNRESFD 180

Query: 214 RFVTRSKIISEIRRFMSNGFLEVETPVLHNEAGGASARPFITHHNAQDIDMVLRIATEL 273
 RFVTRSK+ISEIRR++D FLEVETPVLHNEAGGA+ARPF+THHNAQ+IDMVLRIATEL
 Sbjct: 181 RFVTRSKMISEIRRYLDGLDFLEVETPVLHNEAGGAAARPFVTHHNAQNIDMVLRIATEL 240

Query: 274 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVT 333
 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIE YQAYADY DIM+LTEGIIQH
 Sbjct: 241 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEVYQAYADYLDIMNLTGIIQHAA 300

Query: 334 KTVKGDGPINQYQTEIKINEPFRVHVMVDAVKEITGIDFWKEMTLEEAQALAEKNVPLE 393
 K V+GDGPI+YQGTETI+INEPFRVHVMVDA+KE+TG DFW EMT+EEA ALA+EK VPLE
 Sbjct: 301 KAVRGDGPIDYQGTETIRINEPFRVHVMVDAIKEVTGADFWPEMTVEEAIALAKEKQVPLE 360

Query: 394 KHFTTVGHIINAFEEFVEDTLIQPTFVFGHPVEVSPLAKKNDTPRFTDRFELFIMTKE 453
 KHF +VGHIINAFEEFVE+TL+QPTFVFGHPVEVSPLAKKN D RFTDRFELFIMTKE
 Sbjct: 361 KHFIISVGHIINAFEEFVEETLVQPTFVFGHPVEVSPLAKKNPEDTRFTDRFELFIMTKE 420

Query: 454 YANAFTELNDPIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGMPPPTGGLGIGIDRL 513
 YANAFTELNDPIDQLSRFEAQ AKELGDDEATG+DYD+VEALEYGMPPPTGGLGIGIDRL
 Sbjct: 421 YANAFTELNDPIDQLSRFEAQAKELGDDEATGIDYDFVEALEYGMPPPTGGLGIGIDRL 480

Query: 514 CMLLTDTTTTIRDVLLFPIMK 533
 CMLLT+TTTIRDVLLFPIMK
 Sbjct: 481 CMLLTNTTTTIRDVLLFPIMK 500

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-421-

Example 331

A DNA sequence (GBSx0361) was identified in *S.agalactiae* <SEQ ID 1075> which encodes the amino acid sequence <SEQ ID 1076>. This protein is predicted to be 6,7-dimethyl-8-ribityllumazine synthase (ribH). Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.1042(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:CAB14257 GB:Z99116 riboflavin synthase (beta subunit) [Bacillus subtilis]
      Identities = 103/151 (68%), Positives = 120/151 (79%)

      Query: 1   MTIEGQLVANEMKIGIVVSREFNELITSKLLSGAUGLLRHGVSEEDIDIVWVPGAFEP 60
                M II+G LV +KIGIVV RFN+ ITSKLLSGA D LLRHGV DID+ WVPGAFEP
      Sbjct: 1   MNIIQQNLVGTGLKIGIVVGRFNDFITSKLLSGAEDALLRHGVDTNDIDVAVWVPGAFEP 60

20   Query: 61   YMARKMALYKDYDAIICLGVIKGSTHDYDYVCNEVTGIGHLNSQSDIPHIFGVLTTDN 120
                + A+KMA K YDAII LG VI+G+T HYDYVCNE KGI + + +P IFG++TT+N
      Sbjct: 61   FFAKKMAETKKYDAIITLGTVIRGATTHYDYVCNEAAKGIAQAANTTGVVPVIFGIVTTEN 120

25   Query: 121  IEQAIERAGTKAGNKGVDCAISAIEMVNLDK 151
                IEQAIERAGTKAGNKG DCA+SAIEM NL++
      Sbjct: 121  IEQAIERAGTKAGNKGVDCAVSAIEMANLNR 151

```

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 332

A DNA sequence (GBSx0362) was identified in *S.agalactiae* <SEQ ID 1077> which encodes the amino acid sequence <SEQ ID 1078>. This protein is predicted to be GTP cyclohydrolase ii (ribA/B). Analysis of this protein sequence reveals the following:

```

35   Possible site: 20
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.1918(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9725> which encodes amino acid sequence <SEQ ID 9726> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

50   >GP:AAA86524 GB:U27202 GTP cyclohydrase II/
      3,4-dihydroxy-2-butanone-4-phosphate synthase
      [Actinobacillus pleuropneumoniae]
      Identities = 230/395 (58%), Positives = 307/395 (77%)

      Query: 19   FSPIKKLLQDIKSGKMVLMDDENRENEGDICAEMVTKESINFMAKFGKGLICLPLSN 78
                FS ++ ++ I+ GK++++ DDE+RENEGD ICAAE T E+INFMA +GKGLIC P+S
      Sbjct: 6   FSKVEDAIEAIRQKIIILVTDDRENEGDFICAAEFATPENINFMATYKGLICTPIST 65

```

-422-

Query: 79 YYAEKLELAQMASHNTDNHETAFTISIDHLSTSTGISAE DRALTAKMVANDSSKAKDFRR 138
 A+KL M + N DNHETAFT+S+DH+ T TGISA +R++TA + +D++KA DFRR
 Sbjct: 66 BIAKKLNHFMVAVNQDNHETAFTVSDHIDTGTGISAFERSITAMKIVDDNAKATDFRR 125

Query: 139 PGHLPFLLAKEGGVLRNGHTEATVDLCRLAGLKECGLCCEIMAEDGSMRKDELLAFAQ 198
 PGH+FPL+AKEGGVL RNGTHTEATVDL RLAGLK GLCCEIMA+DG+MM +L FA
 Sbjct: 126 PGHMFPLIAKEGGVLRNGHTEATVDLARLAGLKHAGLCCEIMADDGTMMPDLQKFAV 185

Query: 199 KHDLAIAITIKQLQDYRRQEEGGVREIEIQLPQFGHFTAYGYSEVVANKEHVALVKGDI 258
 +H++ TI+QLQ+YRR+ + V + +++PT++G F A+ + EV++ KEHVALVKGD+
 Sbjct: 186 EHNMPFITIQQLQEYRRKHDSLQISVVKMPTKYGEFMAHSFVEVISGKEHVALVKGDL 245

Query: 259 SSGEDVLCRLHSECLTGDVHSLRCDGCEQLANALQQIEAEGRVLLYMRQEGRGIGLIN 318
 + GE VL R+HSECLTGD F S RCDG+Q A A+ QIE EGRGV+LY+RQEGRGIGLIN
 Sbjct: 246 TDGEQVLARIHSECLTGDAGFSQRCDGQQAAMTQIEQEGRVILYLRQEGRGIGLIN 305

Query: 319 KIKAYHLQEEGLDTLEANLALGFEGDERDYGVSAQLLKDLGINSINLLTNNPDKIQQLA 378
 KL+AY LQ++G+DT+EAN+ALGF+ DER+Y + AQ+ + LG+ SI LLTNNP KI+ L+
 Sbjct: 306 KLRAYELQDKGMDTVEANVALGFKEDEREYYIGAQMFGQLGVKSIRLLTNNPAKIEGLKE 365

Query: 379 EGICVKNRVPLQVAVTAYDLNLYLTKKEKMGHLLD 413
 +G+ + R P+ V D++YLK K+ KMGH+ +
 Sbjct: 366 QGLNIVAREPIIVEPNKNDIDYLVKVKQIKMGHMFN 400

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 333

A DNA sequence (GBSx0363) was identified in *S.agalactiae* <SEQ ID 1079> which encodes the amino acid sequence <SEQ ID 1080>. This protein is predicted to be riboflavin synthase alpha chain (ribE). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3517(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9723> which encodes amino acid sequence <SEQ ID 9724> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05274 GB:AP001512 riboflavin synthase alpha subunit [Bacillus halodurans]
 Identities = 98/216 (45%), Positives = 147/216 (67%), Gaps = 2/216 (0%)

Query: 1 MFTGIIIEEMQVSRIRNGIKSQQLSIDAPKLVPLLRKGDVAVNGVCLTVLDKSETAFIA 60
 MFTGIIIE++G + I+ ++ ++I + K+V ++ GDS+AVNGVCLTV ++T F
 Sbjct: 1 MFTGIIEDVGTIDAIQQTGEAIVMTITSSKKIVSDVQLGDSIAVNGVCLTVTSFTDTQFTV 60

Query: 61 DVPESMMRTSLAALRLHSHKVNLELALRSDSRLGGHVFVLDGVDGKGIEKIQKDDIAVRF 120
 D+MPE++ TSL L S+VNLE A+ ++ R GGH V GHVDG+G I K ++ D AV +
 Sbjct: 61 DLPETVTRATSLRLLSKGSRVNLERAMVANGRFGGHIVSGHVDGIGTIRKKERKDNVAVY 120

Query: 121 SIDAPPSIMSYIIEKGSVALDGISLTVVSFTEHSFEVSVIPHTMAQTNLSLKKVGDLLNI 180
 +I+ S+ Y+I KGSVA+DG SLT+ ++ +F +S+IPHTM +T + LKK GD++NI
 Sbjct: 121 TIEVSSSLRRYMIHKGSVAVDGTSLTIFDVSDKTFTTISIIPHTMEETIIGLKKAGDIVNI 180

-423-

Query: 181 EVDVLGKYAEKFLAPTNRNHTSSVMDWSFLENGY 216
 E D++GKY E+F+ N + +FL+E+GY
 Sbjct: 181 ECDLIGKYIEQFVQQGKPVNEGG--LTKAFLTEHGY 214

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 334

10 A DNA sequence (GBSx0364) was identified in *S.agalactiae* <SEQ ID 1081> which encodes the amino acid sequence <SEQ ID 1082>. This protein is predicted to be riboflavin-specific deaminase (ribD). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.01 Transmembrane 307 - 323 (307 - 323)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA86522 GB:U27202 riboflavin-specific deaminase [Actinobacillus pleuropneumoniae]
 Identities = 182/353 (51%), Positives = 259/353 (72%)
 25 Query: 6 DYMALALKEAEKMGFVAPNPLVGAVIVKDDRIISKGYHKRFGDLHAERQAIKNADEDIS 65
 DYM A+ A++G+G+ PNLVG VIVK+ I+++GYH++ G HAER A+ + ED+S
 Sbjct: 51 DYMRAIALAKQGLGWTNPNPLVGCIVKNGEIVAEGYHEKIGGWHAERNAVHCKEDLS 110
 30 Query: 66 GSTLYVTLEPCCHVGKQPPCTEALIKSGIKKVVVGS LDPNPLVSGKGIALLRKEGLNVEV 125
 G+T VYTLPECC H+ PPC++ LI+ GIKKV +GS DPNPLV+G+G LR+ G+ V
 Sbjct: 111 GATAVYTLPECCHHGRTPPCSDLLIERGIKKVFIGSSDPNPLVAGRGANQLRQAGVEVVE 170
 35 Query: 126 GILREECDALNERFIFHMTYKQPFVYLKYAMTLDGKIATKTGDSKWISNEHSRQSVQKLR 185
 G+L+EECDALN F ++ K+P+V +KYAMT DGKIAT +G+SKWI+ E +R VQ+ R
 Sbjct: 171 GLLKEECDALNPIFFHYIQTKRPVYLMKYAMTADGKIATGSGESKWITGESARARVQQTR 230
 40 Query: 186 QKCSAIMVGINTVLADNPRLTCRIPKGEALVRIVCDSQLKIPLDSYLVKSAKTIPTWIAT 245
 + SAIMVG++TVLADNP L R+P + VRIVCDSQL+ PLD LV++AK T IAT
 Sbjct: 231 HQYS AIMVGVDTVLADNPMLNSRMENAKQPVRIVCDSQLRTPDCQLVQTAKEYRTVIAT 290
 45 Query: 246 CSDNLAQQQTLKEMGCRLIKVPRKDGKLDLKVIMTILGQEGIDSLLEGGSSSLHFSALKA 305
 SD+L + + + +G ++ ++ ++DL+ L+ LG+ IDSLL+EGGSSL+FSAL++
 Sbjct: 291 VSDDLQKIEQFRPLGVDVLVCKARNKRVDLQDLLQKLGEQIDSLLEGGSSSLNFSALLES 350
 Query: 306 GIVNRLIVFIAPKIIGGLKAKTAISGEGLDWLNQAFRVKDIELSRMDSDVVIE 358
 GIVNR+ +IAPK++GG +AKT I GEG+ ++QA ++K + D++++
 Sbjct: 351 GIVNRVHCYIAPKLVGGKQAKTPIGGEGIQIDQAVKLKSTELIGEDILLD 403

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1083> which encodes the amino acid sequence <SEQ ID 1084>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 55 INTEGRAL Likelihood = -1.17 Transmembrane 88 - 104 (88 - 105)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>

-424-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAB11794 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
Identities = 71/161 (44%), Positives = 109/161 (67%)

Query: 13 LEEQTYFMQEALKEAEKSLQKAEIPIGCVIVKDGEEIIGRGNAREESNQAIMHAEMMAIN 72
+ + +M+EA+KEA+K+ +K E+PIG V+V +GEII R HN RE ++I HAEM+ I+

10 Sbjct: 1 MTQDELYMKAEIKAEKAEKGEVPIGAVLVINGETIARAHNLRETEQRSIAHAEMLVID 60

Query: 73 EANAHEGNWRLDITTLFVTIEPCVMCSGAIGLARIPHVYIGASNQKFGGVDSLYQILTDE 132
EA G WRL TL+VT+EPC MC+GA+ L+R+ V++GA + K G +L +L +E

15 Sbjct: 61 EACKALGTWRLEGATLYVTIEPCPMCGAVVLSRVEKVVFGAFDPKGGCSGTLNLLQEE 120

Query: 133 RLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAXHLIKE 173
R NH+ +V G+L +C ++ FFR+ R++KK A+ + E

Sbjct: 121 RFNHQAEVVSGVLEEECGMLSAPFRELRKKKKAARKNLSE 161

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 48/146 (32%), Positives = 71/146 (47%), Gaps = 21/146 (14%)

Query: 7 YMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYHKRFGD---LHAERQAIKNADE 62
+M ALKEAEK + A P +G VIVKD II +G++ R +HAE AI A+

25 Sbjct: 19 FMQEALKEAEKSLQ-KAEIP-IGCVIVKDGEEIIGRGNAREESNQAIMHAEMMAINEANA 76

Query: 63 D-----ISGSTLYVTLEPCCHVGKQPCTEALIKSGIKKVVVGSGLDPNPLVSGKGIALLR 117
+ +TL+VT+EPC C+ A+ + I V+ G+ + +L

30 Sbjct: 77 HEGNWRLDITTLFVTIEPCV-----MCSGAIGLARIPHVYIGASNQKFGGVDSLYQILT 130

Query: 118 KEGLN----VEVGILRECDALNERF 139
E LN VE G+L +C + + F

Sbjct: 131 DERLNHRVQVERGLLAADCANIMQTF 156

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 335

A DNA sequence (GBSx0365) was identified in *S.galactiae* <SEQ ID 1085> which encodes the amino acid sequence <SEQ ID 1086>. This protein is predicted to be Nramp metal ion transporter. Analysis of this

40 protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.89	Transmembrane	169 - 185 (160 - 191)
INTEGRAL	Likelihood = -11.09	Transmembrane	140 - 156 (128 - 165)
45 INTEGRAL	Likelihood = -6.85	Transmembrane	359 - 375 (354 - 379)
INTEGRAL	Likelihood = -6.48	Transmembrane	269 - 285 (263 - 287)
INTEGRAL	Likelihood = -6.16	Transmembrane	426 - 442 (423 - 445)
INTEGRAL	Likelihood = -5.57	Transmembrane	62 - 78 (58 - 80)
INTEGRAL	Likelihood = -4.94	Transmembrane	107 - 123 (103 - 127)
50 INTEGRAL	Likelihood = -4.46	Transmembrane	391 - 407 (389 - 408)
INTEGRAL	Likelihood = -4.35	Transmembrane	310 - 326 (307 - 328)

----- Final Results -----

bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella]

-425-

fastidiosa]

Identities = 192/436 (44%), Positives = 274/436 (62%), Gaps = 14/436 (3%)

Query: 10 SLSEVNQSVVEPHNSSFWNTLRAFLGPGALVAVGYMDPGNWITSVIGGATYRYLLLFVVL 69
 5 SL E++ SV V + L AFLGPG +V+VGYPMDPGNW T + GG+ + Y+LL V+L
 Sbjct: 39 SLGEMHASVAVSRRGHGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFGYMLLSVIL 98

Query: 70 VSSLMAQLQQMAGKLGIVTRQDLAQATASRLPKPLRYLLFIIEELALITDLAEVIGSA 129
 10 +S++MA+ LQ +A +LGI + DLQA +R + L+++ ELA+IA DLAEVIG+A
 Sbjct: 99 LSNVMAIVLQALAAARLGASDMDLAQACRARSRTTALWVVCELATACDLAEVIGTA 158

Query: 130 IALHLLFGWPLLLSIMITILDVFLLLLMKLGKIEAFVSVLILITILIIIFTYLVVLSQP 189
 15 IAL+LL G P++ ++IT +DV L+LLLM G + +EAFV L+L I F +VL+ P
 Sbjct: 159 IALNLLGVPIIWGVVITAVDVVLLVLLMHRGFRALEAFVIALLLVIFGCFVQIVLAAP 218

Query: 190 DLDAMFKGFLPHRELFNISHGKNSPLTLALGIIGATVMPHNLYLHSSLSQTRVDYHNK 249
 L + GF+P ++ L LA+GI+GATVMPHNLYLHSS+ QTR
 Sbjct: 219 PLQEVGGFVPRVQVV----ADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAYP-RTP 272

Query: 250 SSIKKAVERFMTLDSNIQLSLAFVVNSLLLVLGASLYG-HANDISAFSQMYLALSDKTTIT 308
 20 + A+R+ DS + L LA +N+ +L+L A++F+ H D+ Q Y L+
 Sbjct: 273 VGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHDFDVEEIEQAYQLLAPVLGV 332

Query: 309 GAVASSFLSTLFAVALLASGQNSTITGTTLTGQIVMEGFLHFKLPQWLIRLCTRLTLPLI 368
 25 G A TLFA ALLASG NST+T TL GQIVMEGFL +L WL R+ TR L ++P+
 Sbjct: 333 GVAA-----TLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLRVLTGLAIVPV 387

Query: 369 FVIALLVGGEENTLDQLIVYSQVFLSLALPFSIFPLIYFTSQKSIMGEHANAKWNTYLAY 428
 V+ L G E +L++ SQV LS+ LEF++ PL+ + + +MG +W +A+
 30 Sbjct: 388 IVVVALYG--EQGTGRLLLSQVILSMQLEFAVIPLLRVADRVKMGALVAPRWLMVVAW 445

Query: 429 LVAIILTLLNLKLIMD 444
 L+A ++ +LN+KL+ D
 Sbjct: 446 LIAGVIVVLNVKLLGD 461

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 336

40 A DNA sequence (GBSx0366) was identified in *S.agalactiae* <SEQ ID 1087> which encodes the amino acid sequence <SEQ ID 1088>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -14.12 Transmembrane 113 - 129 (98 - 132)
 INTEGRAL Likelihood = -12.15 Transmembrane 228 - 244 (220 - 249)
 INTEGRAL Likelihood = -10.83 Transmembrane 175 - 191 (167 - 195)
 INTEGRAL Likelihood = -5.04 Transmembrane 57 - 73 (55 - 75)
 INTEGRAL Likelihood = -3.93 Transmembrane 146 - 162 (142 - 166)
 50 INTEGRAL Likelihood = -1.38 Transmembrane 199 - 215 (199 - 215)
 INTEGRAL Likelihood = -0.32 Transmembrane 82 - 98 (82 - 98)

----- Final Results -----

bacterial membrane --- Certainty=0.6647(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF11325 GB:AE002018 hypothetical protein [Deinococcus radiodurans]
 Identities = 63/215 (29%), Positives = 108/215 (49%), Gaps = 13/215 (5%)

60 Query: 11 LLLVFILTIIVNYLSATGFLTGNSQKSLSDRYQTLLTPAPLAFSIWSVIYL-LTFLVILR 69

-426-

LL +LT++VNYLS L GNS +SDR TPA L F++W I+L L + +
 Sbjct: 10 LLAATVLTLLVNYLSNALPLFGNSNAEVSDRLPNATPAGLTFTVWGPFLGLLVFAVYQ 69
 Query: 70 AIFSKSQSYQDNFASIFPYFLGLLLVNNIWTVFFTSNLIGLSTIIIFAYCILLV-IIIKI 128
 A+ ++ + D +P+ LG LL N W + F S IGLS +I+ A +LV + + +
 Sbjct: 70 ALPAQRGARLDRL--FWPFLGLNLL-NVAWLLAFQSLNIGLSVVMILALLAVLVRLYLSV 126
 Query: 129 LS---KNKSKLLLRITFGIHAGWLLVASLVNLAVYLVKI---DFNYPLPKVYIAIALI 181
 S + + L++ ++ W+ VA++ N+ +LV F V+ A++ ++
 Sbjct: 127 RSLPPQGAERWTLQLPVSLYLAWISVATIANITAFIVSAGVTQSFLGIAGPVWSALLLV 186
 Query: 182 FITVLSLYLARVLQAYLILSVFWAWLMVFKAHLE 216
 + +L R A+ + + WA+ V+ A E
 Sbjct: 187 AAAIGVFFLWRFRDYAFAAV-LLWAFYGVYVARPE 220

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 337

A DNA sequence (GBSx0367) was identified in *S.agalactiae* <SEQ ID 1089> which encodes the amino acid sequence <SEQ ID 1090>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC65352 GB:AE001215 T. pallidum predicted coding region
 TP0352 [Treponema pallidum]
 Identities = 28/64 (43%), Positives = 41/64 (63%)

Query: 3 EFTFEIVEKLLVLSENEKGWTKELNVRVSPNGAPAKFDLRTWSPDHTKMKGKITLSNEEFK 62
 +F +E+ LS + GW+ EL +S+NG P K+D+R WSPD +KMGKG+TL+ E
 Sbjct: 12 DFHYEVTRNWGTLSTSGNGWSLELKSISWNGRPEKYDIRAWSPOKSKMGKGVTLTRAEIV 71
 Query: 63 VILD 66
 + D
 Sbjct: 72 ALRD 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1091> which encodes the amino acid sequence <SEQ ID 1092>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 59/70 (84%), Positives = 64/70 (91%)
 Query: 1 MSEFTFEIVEKLLVLSENEKGWTKELNVRVSPNGAPAKFDLRTWSPDHTKMKGKITLSNEE 60
 M+EFTF I E LL LSEN+KGWTKELNVRVSPNGA AK+D+RTWSPDHTKMKGKITL+NEE
 Sbjct: 1 MAEFTFNIEEHLTLSENDKGWTKELNVRVSPNGARAKWDIRTWSPDHTKMKGKITLINEE 60

-427-

Query: 61 FKVILDAPRK 70
 FK ILDAFRK
 Sbjct: 61 FKTILDAFRK 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 338

A DNA sequence (GBSx0368) was identified in *S. agalactiae* <SEQ ID 1093> which encodes the amino acid sequence <SEQ ID 1094>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 (92 - 110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14676 GB:Z99117 similar to protease [Bacillus subtilis]
 Identities = 201/407 (49%), Positives = 277/407 (67%), Gaps = 2/407 (0%)

Query: 4 VKKREVLSPAGTLEKIKVAIDYGADAVFVGQAYGLRSRAGNFSMEELQEGINIAHARD 63
 + K+PE+L+PAG LEKIK+A+ YGADAVF+GGQ YGLRS A NF++EE+ EG+ +A
 Sbjct: 18 ITKKPELLAPAGNLEKIKIAVHYGADAVFIGGQEYGLRSNADNFTIEELAEQVEFAKKYG 77

Query: 64 AKVYVAANMVTHEGNELGAGPWFRELDMGLDAVIVSDPALIVICATEAPGLEIHLSTQA 123
 AK+YV N+ H N G + + L D + +IV+DP +I C AP +E+HLSTQ
 Sbjct: 78 AKIYVTTNIFAHNENMDGLELYKALGDANVAGIIVADPLIETCRRVAPNVEVHLSTQQ 137

Query: 124 SSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLS 183
 S +N++ +FWKE GL RVVLARE + E+ E++++ D+EIE+F+HGAMCI+YSGRCVLS
 Sbjct: 138 SLSNWKAVQFWKEEGLDRVVLARETSALREIREMKEKVDIEIESFIHGAMCIAYSRCVLS 197

Query: 184 NHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFMSAVDMCMIEHIPDMI 243
 NHM+ RD+NRGGC QSCRW YDLY G +L GE PF+MS D+ +IE IP MI
 Sbjct: 198 NHMTARDSNRGGCCQSCRWDYDLYQTD-GANAVLYGEEDAPFAMSPKDLKLIESIPKMI 256

Query: 244 ENGVDLSLKIEGRMKSIIHYVSTVTNCKYKAAVDAYMESPEAFEAIKEDLIDELWKVAQRELA 303
 E G+DSLKIEGRMKSIIHYV+TV + Y+ +DAY PE F I+++ ++EL K A R+ A
 Sbjct: 257 EMGIDLSLKIEGRMKSIIHYVATVVSRYRKVIDAYCADPENF-VIQKEWLEELDKCANRDTA 315

Query: 304 TGFYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQNVIMEGDRVEFYGPG 363
 T F+ TP EQ+FG K Y FVG V+++D T++QRN +GD VEF+GP
 Sbjct: 316 TAFEGTTPGYEQMFGEHAKKTTYDFVGLVLNYDEDTQMVTLQQRNFFKKGDEVEFFGPE 375

Query: 364 FRHFECFIDGLRDAEGNKIDRAPNPMELLTITLNPVKKGDMIRACK 410
 +F I+ + D +GN++D A +P++++ L + +M+R K
 Sbjct: 376 IENFTHTIETIWDGDNELDAARHPLQIVKFKLDKKIYPSNMMRKKG 422

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1095> which encodes the amino acid sequence <SEQ ID 1096>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 (92 - 110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-428-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 5 >GP:BAB04993 GB:AP001511 protease [Bacillus halodurans]
Identities = 201/403 (49%), Positives = 280/403 (68%), Gaps = 4/403 (0%)
- Query: 6 KRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAHARGAK 65
K+PE+L+PAG+LEKLKVAI YGADAV++GGQ +GLRS A NFS+EE++EG+++A+ GAK
Sbjct: 17 KKPELLAPAGSLEKLKVAIHYGADAVYIGGQEFGLRSNADNFSIEEMREGVEFANKYGAK 76
- 10 Query: 66 VYVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASS 125
VYV N+ H N G E+ L+++G+ +IV+DP +I C AP +E+HLSTQ S
Sbjct: 77 VYVTINIAHNENMDGLEEYLSALQEVGVGTGIIVADPLIETCKRVAPKVEVHLSTQQL 136
- 15 Query: 126 TNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNH 185
+N+ +FWK GL RVVLAREV + E+ E++K D+EIE FVHGAMCISYSGRCVLSNH
Sbjct: 137 SNWLAVKFWKEGLHRVVLAREVGLLEEMLEMKKHVDIEIETFVHGAMCISYSGRCVLSNH 196
- 20 Query: 186 MSHRDANRGGCSQSCRWKYDLYDMPFGGE-RRSLKGEIPEDYSSVDMCMIDHIPDLIE 244
M+ RD+NRGGC QSCRW YDLY+ E +G++P Y+MS D+ +I IP LIE
Sbjct: 197 MTARDENRGGCCQSCRWDYDLYEQQDSABEIPFAEGDVP--YTMSPKDLNLIQAIPQLIE 254
- 25 Query: 245 NGVDSLKIEGRMKSIHYVSTVTNICYKAAVGAYMESPEAFYAIKEELIDELWKVAQRELAT 304
G+DSLK+EGRMKSIHYV+TVT+ Y+ +AY P+ F IK E ++EL K A R+ A
Sbjct: 255 AGIDSLKVEGRMKSIHYVATVTSVVRKVIDAYCSDPDFNF-KIKREWLEELEKCANRDFAP 313
- 30 Query: 305 GFYYGIPTENEQLFGARRKIPQYKFVGEVVFDSASMTATIRQNRVIMEGDRIECYGPFG 364
F+ G PT EQ++G K +Y FVG V+ ++ + T++QRN +GD +E +GP
Sbjct: 314 QFFEFTPTYKEQMYGIHPKRTKYDFVGLVLDYNEKTGIVTLQRNHFQKGDEVEFFGPEI 373
- 35 Query: 365 RHIFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIR 407
F V+ + D DG ++D A +P++++ + ++V P +M+R
Sbjct: 374 NRFTQIVEKIWDEDCNELDAARHPLQIVKFKVDQKVYPQNMNR 416
- 35 An alignment of the GAS and GBS proteins is shown below:
Identities = 386/427 (90%), Positives = 404/427 (94%)
- Query: 1 MSNVKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGINIAH 60
MS++KKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGI+YAH
40 Sbjct: 1 MSHMKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAH 60
- Query: 61 ARDAKVYVAANMVTHEGNEIGAGPWFRELDMGLDAVIVSDPALIVICATEAPGLEIHL 120
AR AKVYVAANMVTHEGNE+GAG WFR+LRDMGLDAVIVSDPALIVIC+TEAPGLEIHL
Sbjct: 61 ARGAKVYVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHL 120
- 45 Query: 121 TQASSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180
TQASSTNYETFEFWK MGLTRVVLAREV MAELAEIRKRTDVEIEAFVHGAMCISYSGRC
Sbjct: 121 TQASSTNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180
- 50 Query: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFMSAVDMCMIEHIP 240
VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFG ER+SLKGEIPE +SMS+VDMCMI+HIP
Sbjct: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGGERRSLKGEIPEDYSSVDMCMIDHIP 240
- 55 Query: 241 DMIENGVDLSKIEGRMKSIHYVSTVTNICYKAAVDAYMESPEAFEAKEELIDELWKVAQR 300
D+IENGVDLSKIEGRMKSIHYVSTVTNICYKAAV AYMESPEAF AIKE+LIDELWKVAQR
Sbjct: 241 DLIENGVDLSKIEGRMKSIHYVSTVTNICYKAAVGAYMESPEAFYAIKEELIDELWKVAQR 300
- 60 Query: 301 ELATGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQNRVIMEGDRVEFY 360
ELATGFYY PTENEQLFGARRKIPQYKFVGEVV+FD+A M ATIRQNRVIMEGDR+E Y
Sbjct: 301 ELATGFYYGIPTENEQLFGARRKIPQYKFVGEVVFDSASMTATIRQNRVIMEGDRIECY 360
- 65 Query: 361 GPGFRHFECFIDGLRDAEGNKIDRAPNPMELLTITLFPVKKGDMIRACKEGLVNLYQND 420
GPGFRHFE + L DA+G KIDRAPNPMELLTI+LP VK GDMIRACKEGLVNLYQ D
Sbjct: 361 GPGFRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLYQKD 420
- Query: 421 GTSKTVR 427

GTSKTVR

Sbjct: 421 GTSKTVR 427

SEQ ID 1094 (GBS385) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 7; MW 75.7kDa).

The GBS385-GST fusion product was purified (Figure 213, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 312), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 339

A DNA sequence (GBSx0369) was identified in *S.galactiae* <SEQ ID 1097> which encodes the amino acid sequence <SEQ ID 1098>. This protein is predicted to be collagenase. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2208(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14677 GB:Z99117 similar to protease [Bacillus subtilis]

Identities = 92/304 (30%), Positives = 161/304 (52%), Gaps = 5/304 (1%)

Query: 1 MEKIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDLREIAKLVDHAGKE 60

M+K L T S + L+ G VGE+ YGLRL FS +++ + ++ H G +

Sbjct: 1 MKKPELLVTPTSTADILPLIQAGATAFLVGEQRYGLRLAGEFSREDVTKAVEIAHKEGAK 60

Query: 61 LTVACNALMHQEMMDNIKPFLELMKEINVDYLVVGDAVGIFYINKRDGYNFKLIYDTSVVF 120

+ VA NA+ H + + + +L + E VD V GD V + + KL + T

Sbjct: 61 VYVAVNAIFHNDKVGELGEYLAFLAEAGVDAAVFGDPAVLMAARESAPDLKLHWSTETTG 120

Query: 121 TSSRQVNFQGHGAVETVLAREIPSEELFKMSENLEFPAILVYGASVIHHSKRPLLQNY 180

T+ N+WG+ GA +VLARE+ + + ++ EN E EI V+G + + SKR L+ NY

Sbjct: 121 TNYTCNYWGRKGAARSVLARELNMDISIVEIKENAEVEIRIQVHGTMCFQSKRSLIGNY 180

Query: 181 YNF---THITDEKTRERGLFLAEPGDPESHYSIYEDKHGTHIFINNDINMMTKVTELVEH 237

+ + + K +E G+FL + + ++ Y I+ED++GTHI ND+ ++ ++ EL++

Sbjct: 181 FEYQGVMDIERKKKESGMFLHDK-ERDNKYPIFEDENGTHIMSPNDVCIIDELERLIDA 239

Query: 238 HFTHWKLDGIYCPGDNFVAIAEIVETARL-IENTFTQDQAFLEIRKLPKGRGLD 296

+K+DG+ + + + +++ E L +EN + + + ERI + P R +D

Sbjct: 240 GIDSFKIDGVLKMPYLLIEVTKMYREALDLCVENRDEYEAKKEDWIERIESIQPVNRKID 299

Query: 297 TGFY 300

TGF+

Sbjct: 300 TGFF 303

A related GBS nucleic acid sequence <SEQ ID 10949> which encodes amino acid sequence <SEQ ID 10950> was also identified.

-430-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1099> which encodes the amino acid sequence <SEQ ID 1100>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1716(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 245/308 (79%), Positives = 273/308 (88%)

Query: 1 MEKIIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDELREIAKLVDHAGKE 60

MEKII+TATAESIEQVK LLA G+DRIVVGE NYGLRLPH+FS DELR+IAKLVDHAGKE

Sbjct: 1 MEKIIITATAESIEQVKALLAAGVDRIYVGEANYGLRLPHNFSYDELRIAKLVHDAGKE 60

Query: 61 LTVACNALMHQEMMDNIKPFLELMKEINVLYLVVG DAGVFYINKRDGYNFKLIYDTSVVF 120

LTVACNALMHQ+MMD IKPFL+LM EI VDYLTVVG DAGVFY+NKRDGYNFKLIYDTSVVF

Sbjct: 61 LTVACNALMHQDMMDQIKPFLDLMI EIAVDYLTVVG DAGVFYVKNKRDGYNFKLIYDTSVVF 120

Query: 121 TSSRQVNFQWQHGA VETVLAREIPSEELFKMSENLEFP AEI LVYGASVIHHSKRPLLQNY 180

TSSRQVNFQWQHGA VE+VLAREIPS ELF ++ENLEFP AE+LVYGASVIHHSKRPLL+NY

Sbjct: 121 TSSRQVNFQWQHGA VESVLAREIPS AE LFTLAENLEFP AEVLVGASVIHHSKRPLLENY 180

Query: 181 YNFTHTID EKTRERGLFLAEPGD PRESHYSIYEDKHGTHIFINNDINMMTKVTELVEHHFT 240

Y+FT I DE +RERGLFLAEPGD SHYSIYED HGTHIFINNDI+MM+K+ EL H T

Sbjct: 181 YHFTKIDDEVSRERGLFLAEPGD ASSHYSIYEDNHGTHIFINNDIDMMSKLGELYAHGLT 240

Query: 241 HWKLDGIYCPGD NFVAIAE I FVETARLIENGTF TQDQAF LFDERIRKLHPKGRGLDTGFY 300

HWKLDGIYCPGD+FVAI ++F++ L+E G FTQ++A D+ + HP GRGLDTGFY

Sbjct: 241 HWKLDGIYCPGD NFVAITKLF IQAKTLL EAGQFTQEAEKLDQAVHAHHPAGRGLDTGFY 300

Query: 301 DFDPSTVK 308

+FDP TVK

Sbjct: 301 EFDPKTVK 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 340

A DNA sequence (GBSx0371) was identified in *S.agalactiae* <SEQ ID 1101> which encodes the amino acid sequence <SEQ ID 1102>. This protein is predicted to be cDNA EST yk542c12.5 comes from this gene. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD15622 GB:U75480 unknown [Streptococcus mutans]

Identities = 69/152 (45%), Positives = 101/152 (66%), Gaps = 12/152 (7%)

Query: 1 MSKLFKTLVISAASGAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60

MSK KT +I A +GAAAYFL+T KGK+ +K + + +YKENP+EYHQ A DK +EY

Sbjct: 1 MSKFLKTAIIGAGTGAAAYFLSTDKGKQFKKKIHQTFTDYKENPKEYHQYAADKVNEYK 60

-431-

Query: 61 NLAVDTFKDYKGFESGELTTEDIVSAVKEKSGEVVDFAVDFVNQAKSKFSDEDTAKKED 120
 ++AV +FKDYK KFE+GELT ++I+S+VKEK+ + FAN ++Q K + T +K +
 Sbjct: 61 DVAVHSFKDYKDKFETGELTKDNIISVKEKASQAGKFANSKLSQVKDHLA--QTVEKAE 118

Query: 121 KAP-----ETKVEDIVIDYKENTEDKE 142
 + +V+DIVIDY+ + K+
 Sbjct: 119 ASTNDAGIPLGEMKAQVDDIVIDYQAEETKK 150

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1103> which encodes the amino acid sequence <SEQ ID 1104>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.81 Transmembrane 15 - 31 (14 - 31)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related sequence was also identified in GAS <SEQ ID 9117> which encodes the amino acid sequence <SEQ ID 9118>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 19
 >>> Seems to have a cleavable N-term signal seq.

25 ----- Final Results -----
 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below:

Identities = 69/140 (49%), Positives = 91/140 (64%), Gaps = 8/140 (5%)

Query: 1 MSKLFKTLVISAASGAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60
 M+K FK LVI A SG AAAYFL+T+KGK L+ AEK Y YKE+P++YHQ AK+K SEYS
 Sbjct: 8 MNKSFINLVIGAVSGVAAAYFLSTKKGKALKNRAEKAYQAYKESPDYHQFAKEKGSEYS 67

Query: 61 NLAVDTFKDYKGFESGELTTEDIVSAVKEKSGEVVDFAVDFVNQAKSKFSDEDTAKKE 119
 +LA DTF D K K SG+LT ED++ +K+K+ FV + K ++ E K++
 Sbjct: 68 HLAARDTFYDVKDKLASGDLTKEDMLDLLKDKT-----TAFVQRTKETLAEVEAKEKQD 120

Query: 120 DKAPETKVEDIVIDYKENTE 139
 D + EDI+IDY E E
 Sbjct: 121 DVIIDLNEEDIIIDYTEQDE 140

45 SEQ ID 1102 (GBS164) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 4; MW 17.4kDa).

The GBS164-His fusion product was purified (Figure 115A; see also Figure 200, lane 4) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 115B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 341

A DNA sequence (GBSx0372) was identified in *S.agalactiae* <SEQ ID 1105> which encodes the amino acid sequence <SEQ ID 1106>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -16.93    Transmembrane    6 - 22 ( 1 - 31)

----- Final Results -----
10      bacterial membrane --- Certainty=0.7771(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
    Identities = 88/129 (68%), Positives = 112/129 (86%)

Query: 1  MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDEVNVTLYQTNEILAKAN 60
          M EIA+LI+AIAF VLV+ ++ +L+K+S+T++E++QT+K+LTSDVNVTLYQTNE+LAKAN
20  Sbjct: 1  MWEIALLIVAIAFAVLVIYLILLRLKISDTVDESRTLKILTSDVNVTLYQTNEILLAKAN 60

Query: 61  VLVDDVNGKVSTIDPLFVAIADLSESVSDNLNQRHIGQKASSATSSVTKAGSALAIGKA 120
          VLV+DVNGKV TIDPLF AIADLS SVSDLN QAR+ G+K +T++V KAG+A GK
25  Sbjct: 61  VLVEDVNGKVETIDPLFTAADLSVSVSDLNQRARYFGKKTRKSTANVGKAGAAAYTFGKV 120

Query: 121 ASKIFRKKG 129
          ASK+FRKKG
30  Sbjct: 121 ASKLFRKKG 129

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1107> which encodes the amino acid sequence <SEQ ID 1108>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
30  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -0.85    Transmembrane    18 - 34 ( 17 - 34)

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

40  >GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
    Identities = 83/128 (64%), Positives = 110/128 (85%)

Query: 6  ISIMIIALAFVALVIFLIIVLKKVSETIDEAKKTIISVLTSDEVNVTLHQINDILAKANILV 65
          I+L+I+A+AF LVI+LI++L+K+S+T+DE+++T+ +LTSDVNVTL+QTN++LAKAN+LV
45  Sbjct: 4  IALLIVAIAFAVLVIYLILLRLKISDTVDESRTLKILTSDVNVTLYQTNEILLAKANVLV 63

Query: 66  EDVNGKVATIDPLFVAIADLSESLSDNLNQRHFGQKATNATGNVSKAGKLALVGKVASK 125
          EDVNGKV TIDPLF AIADLS S+SDLN QAR+FG+K +T NV KAG GKVASK
50  Sbjct: 64  EDVNGKVETIDPLFTAADLSVSVSDLNQRARYFGKKTRKSTANVGKAGAAAYTFGKVASK 123

Query: 126 VFGKKGEK 133
          +F KKG++
55  Sbjct: 124 LFRKKGKQ 131

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 92/131 (70%), Positives = 116/131 (88%)

60  Query: 1  MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDEVNVTLYQTNEILAKAN 60
          ++ I+++IIA+AFV LV+ ++ VLKKVSETI+E K+TI VLTSDVNVTL+QTN+ILAKAN

```

-433-

Sbjct: 3 LVGISLMIIALAFVALVIFLIIVLKKVSETIDEAKKTISVLTSDVNVTLHQTDILAKAN 62

Query: 61 VLVDVNGKVSTIDPLFVAIADLSESVSDLNQARHIGQKASSATSSVTKAGSALAIGKA 120
+LV+DVNGKV+TIDPLFVAIADLSES+SDLN QARH GQKA++AT +V+KAG +GK

5 Sbjct: 63 ILVEDVNGKVATIDPLFVAIADLSESLSDLNSQARHFGQKATNATGNVSKAGKLALVGKV 122

Query: 121 ASKIFRKKGDK 131
ASK+F KKG+K

10 Sbjct: 123 ASKVFGKKGK 133

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 342

A DNA sequence (GBSx0373) was identified in *S.agalactiae* <SEQ ID 1109> which encodes the amino acid sequence <SEQ ID 1110>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.0462 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 343

A DNA sequence (GBSx0374) was identified in *S.agalactiae* <SEQ ID 1111> which encodes the amino acid sequence <SEQ ID 1112>. This protein is predicted to be prolipoprotein diacylglycerol transferase (lgt). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -8.39 Transmembrane 231 - 247 (225 - 251)
INTEGRAL Likelihood = -7.64 Transmembrane 89 - 105 (87 - 107)
INTEGRAL Likelihood = -5.20 Transmembrane 18 - 34 (13 - 36)
INTEGRAL Likelihood = -1.86 Transmembrane 46 - 62 (46 - 64)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4354 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9721> which encodes amino acid sequence <SEQ ID 9722> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
transferase [Streptococcus mutans] (ver 3)
Identities = 184/257 (71%), Positives = 226/257 (87%)

50 Query: 2 MINPVAIRLGPFSSIRWYAICIVSGMLLAVYLAMKEAPRKNISDDILDFILMAFPPLSIVG 61
MINP+AI+LGP +IRWY+ICIV+G++LAVYL ++EAP+KNISDD+LDFIL+AFPL+IVG

-434-

Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

Query: 62 ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAA 121
AR+YYVIF+W YY K+P EI IW+GGIAIYGGL+TGA++L IFSY R+I PIDFLD+AA

5 Sbjct: 61 ARLYYVIFDWDYLLKNPSEIPVIWHGGIAIYGGLTGALVLFIFSYRMKPIDFLDVAA 120

Query: 122 PGVMIAQAIGRWGNFINQEAYGRAVKNLNYVENFIKNQMYIDGAYRVPTFLYESLWNFLG 181
PGVM+AQ+IGRWGNF+NQEAYG+ V LNY+P+FI+ QMYIDG YR PTFLYESLWN LG

10 Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNYLPDFIRKQMYIDGHYRTPPTFLYESLWNLLG 180

Query: 182 FVIIMSIRHRPRTLKQGEVACFYLWVWGCGRFIEGMRTDSLVLGRLVSVQWLSVILVII 241
F+IIM +R RP LK+GEVA FYL+WYG GRF+IEGMRTDSL A LRVSVQWLSV+LV++

Sbjct: 181 FIIIMILRRRNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSVQWLSVLLVVV 240

15 Query: 242 GIVMIIYRRREQHISYY 258
G++++ RRR I YY

Sbjct: 241 GVILMVIRRRNHAIPYY 257

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1113> which encodes the amino acid
20 sequence <SEQ ID 1114>. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.01	Transmembrane	229 - 245 (222 - 249)
INTEGRAL	Likelihood = -6.90	Transmembrane	45 - 61 (40 - 68)
INTEGRAL	Likelihood = -4.41	Transmembrane	17 - 33 (11 - 35)
INTEGRAL	Likelihood = -4.14	Transmembrane	87 - 103 (86 - 106)
INTEGRAL	Likelihood = -0.27	Transmembrane	170 - 186 (170 - 186)

25

----- Final Results -----

30 bacterial membrane --- Certainty=0.3803(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
transferase [Streptococcus mutans] (ver 3)
Identities = 176/258 (68%), Positives = 217/258 (83%)

40 Query: 1 MINPIALKCGPLAIHWYALCILSGLVLAVYLASKKAPKKGISSDAIFDFILIAFPLAIVG 60
MINPIA+K GPL I WY++CI++GL+LAVYL +EAPK I SD + DFILIAFPLAIVG

Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

Query: 61 ARIYYVIFEWYVYVHLDEIIAIWNGGIAIYGGLITGALVLLAYCYNKVLNPIHFLDIAA 120
AR+YYVIF+W YY+K+ EI IW+GGIAIYGGL+TGALVL + Y +++ PI FLD+AA

45 Sbjct: 61 ARLYYVIFDWDYLLKNPSEIPVIWHGGIAIYGGLTGALVLFIFSYRMKPIDFLDVAA 120

Query: 121 PSVMVAQAIGRWGNFINQEAYGKAVSQLNYLPSFIQKQMFIEGSRIPPTFLYESLWNLLG 180
P VM+AQ+IGRWGNF+NQEAYGK V+QLNYLP FI+KQM+I+G YR PTFLYESLWNLLG

50 Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNYLPDFIRKQMYIDGHYRTPPTFLYESLWNLLG 180

Query: 181 FVIIMMWRKPKSLLDGEIFA FYLIWYGSGRVIEGMRTDSLMFLGIRISQYVSALLIII 240
F+IIM+ RR+P L +GE+ FYLIWYGSGR VIEGMRTDSLMF +R+SQ++S LL+++

Sbjct: 181 FIIIMILRRRNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSVQWLSVLLVVV 240

55 Query: 241 GLIFVIKRRRQKGISYYQ 258
G+I ++ RRR I YYQ

Sbjct: 241 GVILMVIRRRNHAIPYYQ 258

An alignment of the GAS and GBS proteins is shown below:

60 Identities = 176/257 (68%), Positives = 221/257 (85%)

Query: 2 MINPVAIRLGPFSIRWYAICIVSGMLLAVYLAKEAPRKNIKSDDILDFILMAFPLSIVG 61
MINP+A++ GP +I WYA+CI+SG++LAVYLA KEAP+K I SD I DFIL+AFPL+IVG

Sbjct: 1 MINPIALKCGPLAIHWYALCILSGLVLAVYLASKKAPKKGISSDAIFDFILIAFPLAIVG 60

10

15

20

25

30

35

40

45

50

55

60

678 708 738 768 798 828 858 888
QEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLGFVIIMSIRHRPRTLKQGEVACFYLVWYCGGRPIIEGM
||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
QEAYGKTVTQLNLYLPDFIRKOMYIDGHYRTPFTFLYESLWNLLGFIIIMILRRRPNNLLKEGEVAFFYLWIWYSGRGFVIEGM
150 160 170 180 190 200 210

-436-

```

          918      948      978      1008      1038      1068      1098      1128
RTDSLYLAGLRVSQWLSVILVIIGIVMIYRRREQHISYY*TEEVL**KLLY*LLPLRLLF*F*EYFSF*KKYQKRLRKP
||||| :| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5 RTDSLMFASLRVSQWLSVLLVVGVILMVIRRRNHAIPYYQC
          230      240      250

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 344

- 10 A DNA sequence (GBSx0375) was identified in *S.agalactiae* <SEQ ID 1115> which encodes the amino acid sequence <SEQ ID 1116>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have no N-terminal signal sequence

```

- 15 ----- Final Results -----
- ```

bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

- 20 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAA77782 GB:AB027460 Hpr kinase [Streptococcus bovis]
Identities = 264/309 (85%), Positives = 292/309 (94%)

```

- 25 Query: 1 MAVTVQMLVDRLKLNVIYGDDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60  
M+VTV+MLVD++KL+VIYGD+ LLSK ITT+DISRPGLEMTGYFDYY+PERLQL+GMKEW  
Sbjct: 1 MSVTVKMLVDKVKLDVIYGDDEHLLSKRITTADISRPGLEMTGYFDYYSPERLQLLGMKEW 60
- 30 Query: 61 SYLMAMTGHNRVQVLRMFQKETPAIVVARDLEIPEEMYEAAKDTGIAILQSKAPTSRLS 120  
SYL MT HNR VLREM + ETPAI+VAR+L IPEEM AAK+ GIAILQS PTSRLS  
Sbjct: 61 SYLTGMTSHNRHVLREMIKPTPAIIVARNLAIPEEMISAAKEKGIAILQSHVPTSRLS 120
- 35 Query: 121 GEVSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDGSGIGKSETGLELVKRGHRLVADDRVD 180  
GE+SWYLDSCLAERTSVHGVLMDIYGMGVLIQGDGSGIGKSETGLELVKRGHRLVADDRVD  
Sbjct: 121 GEMSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDGSGIGKSETGLELVKRGHRLVADDRVD 180
- 40 Query: 181 VYAKDEETLWGEPAEILRHLEIRGVGIIIDMSLYGASAVKDSSQVQLAIYLENFETGKV 240  
V+AKDEETLWGEPAEILRHLEIRGVGIIID+MSLYGASAVKDSSQVQLAIYLEN+E+GKV  
Sbjct: 181 VFAKDEETLWGEPAEILRHLEIRGVGIIIDVMSLYGASAVKDSSQVQLAIYLENYESGKV 240
- 45 Query: 241 FDLRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKQMGEFDTQTFEDRLT 300  
FDRLGNGNEE+ELSGVK+PR+RIPV+TGRN+SVVIEAAAMN+RAKQMGEFDT+TFE+RLT  
Sbjct: 241 FDLRLGNGNEEIELSGVKIPRLRIPVQTGRNMSVVIEAAAMNYRAKQMGEFDTKTTFEERLT 300
- Query: 301 HLISQNEVN 309  
LI++NE N  
Sbjct: 301 QLITKNEGN 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1117> which encodes the amino acid sequence <SEQ ID 1118>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 13  
>>> Seems to have no N-terminal signal sequence
- 55 ----- Final Results -----
- ```

bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 255/309 (82%), Positives = 288/309 (92%)

```

-437-

Query: 1 MAVTVQMLVDRLKLNVIYGDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60
 M VTV+MLV ++KL+V+Y ++LLSK ITT+DISRPGLEMTGYFDYYAPERLQL GMKEW
 Sbjct: 32 MTVTVKMLVQKVKLDVYATDNLLSKEITTSDISRPGLEMTGYFDYYAPERLQLFGMKEW 91

Query: 61 SYLMAMTGHNRVQVLRMFQKETPAIVVARDLEIPEEMYEAAKDTGIAILQSKAPTSRLS 120
 SYL MT HNRY VL+EMF+K+TPA+VV+R+L IP+EM +AAK+ GI++L S+ TSRL+
 Sbjct: 92 SYLTQMTSHNRYSVLKEMFKDTPAVVVSRLAIPKEMVQAAKEGISLLSSRVSTSLA 151

Query: 121 GEVSWYLDSCLAERTSVHGVLMIDIYGMGVLIQGDGSGICKSETGLELVKRGHRLVADDRVD 180
 GE+S++LD+ LAERTSVHGVLMIDIYGMGVLIQGDGSGICKSETGLELVKRGHRLVADDRVD
 Sbjct: 152 GEMSYFLDASLAERTSVHGVLMIDIYGMGVLIQGDGSGICKSETGLELVKRGHRLVADDRVD 211

Query: 181 VYAKDEETLWGEPAEILRHLLLEIRGVGIIDIMSLYGASAVKDSSQVQLAIYLENFETGKV 240
 VYAKDEETLWGEPAEILRHLLLEIRGVGIID+MSLYGASAVKDSSQVQLAIYLENFE GKV
 Sbjct: 212 VYAKDEETLWGEPAEILRHLLLEIRGVGIIDVMSLYGASAVKDSSQVQLAIYLENFEAGKV 271

Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKMGFDATQTFEDRLT 300
 FDRLGNGNEEI SGV++PRIRIPVKTGRNVSVVIEAAAMNHRAK+MGFDAT+TFEDRLT
 Sbjct: 272 FDRLGNGNEEITFSGVRIPRIRIPVKTGRNVSVVIEAAAMNHRAKMGFDATKTFEDRLT 331

Query: 301 HLISQNEVN 309
 LI++NEV+
 Sbjct: 332 QLITKNEVS 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 345

A DNA sequence (GBSx0376) was identified in *S.agalactiae* <SEQ ID 1119> which encodes the amino acid sequence <SEQ ID 1120>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1836 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9719> which encodes amino acid sequence <SEQ ID 9720> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 346

A DNA sequence (GBSx0377) was identified in *S.agalactiae* <SEQ ID 1121> which encodes the amino acid sequence <SEQ ID 1122>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -4.88 Transmembrane 35 - 51 (31 - 59)

----- Final Results -----

bacterial membrane --- Certainty=0.2954 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-438-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAC67275 GB:AF017113 Yv1C [Bacillus subtilis]
 Identities = 21/63 (33%), Positives = 36/63 (56%), Gaps = 2/63 (3%)

Query: 3 SSFYKQRKGLVCGVVAGLADKYNWDLALSRVLIALILYFTKF--GLLLYILLAVFLPYK 60
 + Y+ K K + GV+ GLA+ +NWD +L RV+ ++ T LL+YI+ +P +
 10 Sbjct: 2 NKLYRSEKNKKIAGVIGGLAEYFNWDASLLRVITVILAIMTSVLPVLLIYIIWIFTVPSE 61

Query: 61 EDI 63
 D+
 Sbjct: 62 RDM 64

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1123> which encodes the amino acid sequence <SEQ ID 1124>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 20 INTEGRAL Likelihood = -5.26 Transmembrane 39 - 55 (31 - 61)

----- Final Results -----
 bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 60/90 (66%), Positives = 77/90 (84%), Gaps = 3/90 (3%)

30 Query: 1 MKSSFYKQRKGLVCGVVAGLADKYNWDLALSRVLIALILYFTKFGLLLYILLAVFLPYK 60
 +++ FYKQRK +LV GV+AGLADKY WDLAL+RVL AL++Y T FG+LLYILLA+FLPYK
 Sbjct: 1 VETKFKYQRKNRLVAGVIAGLADKYGWDLALARVLAALLIYGTGFGVLLYILLAIFLPYK 60

Query: 61 EDIIETR-RQGPRRRKDAEPV--DDDGWFW 87
 ED++E R +GPRRRKDA+ + ++DGWFW
 35 Sbjct: 61 EDLLEERYGRGPRRRKDADVLNEEDGWFW 90

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 347

40 A DNA sequence (GBSx0378) was identified in *S.agalactiae* <SEQ ID 1125> which encodes the amino acid sequence <SEQ ID 1126>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9717> which encodes amino acid sequence <SEQ ID 9718> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

55 >GP:BAE04250 GB:AP001508 unknown conserved protein [Bacillus halodurans]
 Identities = 379/729 (51%), Positives = 515/729 (69%), Gaps = 25/729 (3%)

Query: 29 ENLNITQIAIDLGIKASQIEKVLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLKDS 88

-439-

E I +A +L +K + I++V++L EGNT+PFIARYRKE+TG +DEV+I+ + +
 Sbjct: 8 EEHTIKTLAKELSLKPNYIKQVIQLLHEGNTVPFIARYRKELTGGMDEVKIREVSEKWTY 67
 Query: 89 MTALSDRKTIVLAKIEEQGLTQELKKAIEEATKLADVEELYLPYKEKRRTKATTAREAG 148
 L +RK V+ +EEQGKLT E KK +E+A KL +VE+LY PYK+KRRT+AT+A+E G
 Sbjct: 68 ANQLHERKEEVIRLVEEQGLTDEWKKTVEQAQKLQEVEDLYRPPYKQKRRTATVAKEKG 127
 Query: 149 LFPLARLI--LQKDNLEEEAQNLYLTDGFETTT--KALSGAVDILIEAFSEDNKLRSWTY 204
 L PLA + L + +EA+ YL+ E T L GA DI+ E ++D LR
 Sbjct: 128 LEPLAEWLFSLPRDGPLQEAQVYLSVEHELTKEVDVLOGAQDIIAEWIADADLRKRIR 187
 Query: 205 NEIWNYSITAVVKDESLDEKQVFKIYYDFSEKISKLGHYQVLAALNRGEKMGVLKVNFEH 264
 + + S+ A VK E LDEK V+++YVD+ E + L ++ LALNRGEK VL+V
 Sbjct: 188 SLGFKEGSVIAKVKEELDEKGVYEMYDYEBPVRTLVPHTLALNRGEKEDVLRVTIRF 247
 Query: 265 NLEKMFRR----FAVRFKETS-QYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISL 319
 ++++ F RF + Y+ I K+ I P++ER IR EL+E AE+ AI +
 Sbjct: 248 PVDRIEMSEKTFIRRFSPAVPVYKAAIEDGYKRLIEPSIEREIRHELTEKABEQAIHI 307
 Query: 320 FSENLRLNLLLVSPKGMVLGFDPAFRGTAKLAVVDQTKLMTTQVIYVPVPPANQAKIEQ 379
 F+ENLR+LLL P+KKG+VLG DPA+RTG KLA+VD+TGK++ QVIYP PP N+ +
 Sbjct: 308 FAENLRSLLLQPPKGVVLGLDPAYRTGCKLAIVDETGKVLDTQVIYTPPKNE--VAA 365
 Query: 380 SKIELAKLIKEFNIETIAIGNGTASRESEAFVAEVLQDFPD-VSYVIVNESGASVYSASE 438
 +K + KLI ++ +E+IAIGNGTASRESE F+A++++D P + Y+IVNE+GASVYSASE
 Sbjct: 366 AKKIVKKLIADYGVEIMIAIGNGTASRESEQFIADLIKDLPTIYLLIVNEAGASVYSASE 425
 Query: 439 LARHEFPDLTVEKRSASISARRLQDPLAELVKIDPKSIGVGQYQHDVSQKLAENLDFVV 498
 + R EFPDL VE+RSA+SIARRLQDPLAELVKIDPKS+GVGQYQHDVSQK+L E+L FVV
 Sbjct: 426 IGREEFPDLQVEERSAVSIARRLQDPLAELVKIDPKSVGVGQYQHDVSQKRLNESITFVV 485
 Query: 499 ETVVNQVGVMVNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKKVPRLGAKAF 558
 ETVVNQVGVMVNTASP+LL +V+GL+KT+++NIVK REE G+ +RA++K +PRLGAK +
 Sbjct: 486 ETVVNQVGVMVNTASPSLLQYVAGLSKTVAKNIVKKREEAGRFTARAQLKDIPRLGAKTY 545
 Query: 559 EQAAGFLRIPNAKNFLDNTGVHPESYEA VKLLDQLTIKELD---DLAKEKLQNLDLIAT 615
 EQ GFLRI + N LD T +HPESY+ KLL ++ D + K+KLQ LD+ A
 Sbjct: 546 EQCIGFLRIMDGNLLDATAIHPEYKVTDKLLSEVGATAADVGIEDLKKKLQALDVSAM 605
 Query: 616 AESIGVGQETLKDIIEDLLKPGDRDLRDDFEAPVLRHVDVLDVSDLVKGQELQGTVRNVVDF 675
 A ++ VG TLKD+I+ L++P RD RD+ P+L+ DVL + DL G ELQGTVRNVVDF
 Sbjct: 606 AATLDVGVFTLKDMDALIRPTRDPRDEVAKPLKQDVLQLEDLLPGMELQGTVRNVVDF 665
 Query: 676 GAFVDIGVHEDGLIHQSRLIKRKDKKTRKMPPLQHPSKYLSVGDIVTVWVVEVDAERSR 735
 G FVDIGV +DGL+H S+L R ++HP + ++VG+IVTVWV +VD ++ R
 Sbjct: 666 GVFDIGVQDGLVHISKLANRY-----IKHPLEVTVGEIVTVWVEDVDIKKGR 715
 Query: 736 IGLSLIKPD 744
 I L++++P+
 Sbjct: 716 IALTMLRPE 724

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1127> which encodes the amino acid sequence <SEQ ID 1128>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2207(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 532/716 (74%), Positives = 619/716 (86%), Gaps = 10/716 (1%)
 Query: 28 MENLNITQIAIDLGIKASQIEKVLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLDK 87

-440-

```

MEN N   IA L +   QIE+VL LT +GNTIPFIARYRKE+TGNLDEV IKSIIID+DK
Sbjct: 1  MENNNHNHIAEALSVSLHQIEQVLALTAQGNITIPFIARYRKEVTGNLDEVVIKSIIIDMDK 60

Query: 88  SMTALSDRKTTVLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKPKRRRTKATIAREA 147
5         S+T L++RK T+LAKIEEQGKLT +L+ +IE   KLAD+EELYLPYKPKRRRTKATIAREA
Sbjct: 61  SLTTLNERKATILAKIEEQGKLTDLRTSIEATEKLADLEELYLPYKPKRRRTKATIAREA 120

Query: 148  GLFPLARLILQNKDNLEEEAQNLYLTDGFETTTKALS GAVDILIEAFSEDNKLRSWTYNEI 207
10        GLFPLARLILQN NLE A+ ++T+GF +   +AL+GAVDIL+EA SED KLRSWTYNEI
Sbjct: 121 GLFPLARLILQNAQNLETAEPFVTEGFAS PQEALAGAVDILVEAMSEDAKLRSWTYNEI 180

Query: 208  WNYSSITAVVKDES LDEKQVFKIYYDFSEKISKLHGYQVLALNRGEKMGVLKVNFEHNLE 267
15        W YS + + +KDE LDEK+VF+IYYDFS+++S + GY+ LALNRGEK+G+LKV+FEHNLE
Sbjct: 181 WQYSRLVSTLKDEQLDEKVFQIYYDFS DQVSNMQGYRTLALNRGEKLGILKVSFEHNLE 240

Query: 268  KMFRFFFAVRFKETSQYIDDLIVQT VKKKIVPAMERRIRTELSEGAEDGAISLFSENLRNL 327
20        KM RFF+VRFKET+ YI+++I QT+KKKIVPAMERR+R+ELS+ AEDGAI LFSENLR+L
Sbjct: 241 KMQRFFSVRFKETNPYIEEVINQTIKKKIVPAMERRVSELSDAEDGAIHLSFENLRHL 300

Query: 328  LLVSPLKGMVLGFDPAFR TGAKLAVVDQTKGLMTTQVIYPVPPANQAKIEQSKIELAKL 387
25        LLVSPLKGMVLGFDPAFR TGAKLA+VDQTKGL+TTQVIYPV PA+Q KI+ +K L +L
Sbjct: 301 LLVSPLKGMVLGFDPAFR TGAKLAIVDQTKGLLTQVIYPVAPASQTKIQAAKETLTQL 360

Query: 388  IKEFNIEIIAIGNGTASRESEAFVAEVLQDFDVS YVIVNESGASVYSASELARHEFPDL 447
30        I+ + I+IIAIGNGTASRESEAFVA+VL+DFF+ SYVIVNESGASVYSASELARHEFPDL
Sbjct: 361 IETYQIDIIAIGNGTASRESEAFVADVLKDFNTSYVIVNESGASVYSASELARHEFPDL 420

Query: 448  TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDV SQKKLAENLDFVTVVNQGV 507
35        TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDV SQKKL+ENL FVV+TVVNQGV
Sbjct: 421 TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDV SQKKLSENLFVVDTTVVNQGV 480

Query: 508  NVNTASPALLAHVSGLNKTI SENIVKYREENGQIKSRAEIKKVPRLGAKAFEQAAGFLRI 567
40        NVNTASP+LLAHVSGLNKTI SENIVKYREENG + SRA+IKKVPRLGAKAFEQAAGFLRI
Sbjct: 481 NVNTASPSLLAHVSGLNKTI SENIVKYREENGALTSRADIKKVPRLGAKAFEQAAGFLRI 540

Query: 568  PNAKNFLDNTGVHPESYEAVKLLDQLTIKELDDLAKELQNLDLIATAESIGVGQETLK 627
45        P AKN LDNTGVHPESY AVK+L L I++LDD AK L + + AE++ +GQETLK
Sbjct: 541 PGAKNILDNTGVHPESYPAVKELFKVLGIQDLDDAAKATLA AVQVPQMAETLAIGQETLK 600

Query: 628  DIIEDLLKPGRDRLRDDFEAPVLRHDVLDVSDLVKGQELQGTVRNVVDFGAFVDIGVHEDG 687
50        DII DLLKPGRDRLRDDFEAP+LR D+LD+ DL++GQ+L+GTVRNVVDFGAFVDIGVHEDG
Sbjct: 601 DIIADLLKPGRDRLRDDFEAPILRQDILDLKDLEIGQKLEGTVRNVVDFGAFVDIGVHEDG 660

Query: 688  LIHQSRSLIKRKDKKTRKMPPLQHP SKYLSVGDIVTVVWVEVDAERSRIGLSLIK 743
45        LIH S + K + HPS+ +SVGD+VTWVW ++D +R ++ LSL+ P
Sbjct: 661 LIHISEMSKTF-----VNHP SQVSVGDLVTWVW SKIDLDRHKVNLSLLPP 706

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 348

A DNA sequence (GBSx0379) was identified in *S.agalactiae* <SEQ ID 1129> which encodes the amino acid sequence <SEQ ID 1130>. This protein is predicted to be N5,N10-methylenetetrahydromethanopterin reductase homolog. Analysis of this protein sequence reveals the following:

```

Possible site: 60
55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4864 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
60          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-441-

>GP:AAB94650 GB:U96107 N5,N10-methylenetetrahydromethanopterin
 reductase homolog [Staphylococcus carnosus]
 Identities = 164/300 (54%), Positives = 217/300 (71%), Gaps = 1/300 (0%)

5 Query: 45 VYGIGEHHRDFAVSAPEIVLAAGAVRTNNIRLSSAVTILSSNDPIRVYQQFSTIDALSN 104
 +YG+GEHR D+AVS P VLAA A T I+LSSAVT+LSS+DP+ VY++F+T+DA+SN
 Sbjct: 1 NYGLGEHRSDYAVSDPVTVLAASLTQRIKLSSAVTVLSSDDPVCVYERFATLDAVSN 60

10 Query: 105 GRAEIMAGRGSFIESFPLFGYDLADYDDL FNEKMDMLLAINSATNLDWKGHILTQTVNERP 164
 GRAEIM GRGSFIESFPLFGYDL DYD LF EK+++L IN + W+G + +
 Sbjct: 61 GRAEIMVGRGSFIESFPLFGYDLDDYDLRFVEKLELLKEINQHEVVIWEGTMRPAIKGLG 120

15 Query: 165 IYPRALQRQLPIWVATGCVNDSTIRIAEQGLPIVYATIGGNPKAFRQLVHIYKEVGS RNG 224
 +YPR+Q ++PIW+ATGG +S+IR AE GLPI YA IGGNPK F++ + IY+ V G
 Sbjct: 121 VYPRAVQDEIPIWLATGGTPESSIRAAEFGLPITYAIIGGNPKRFKRNIAYRAVAESRG 180

20 Query: 225 HKPEQLKVAHWSGWIEEDNQTAIDRYFFPTKQIVDNIAGRPWSEMTKEQYLSVGPE 284
 + + VA HSWG+I + ++ A ++ PTK + IAK R +W T+ + R + E
 Sbjct: 181 YDLADMPVAVHWSGYIADTDEQAQREFYEPTKVHREIIAKER-NWPPYTEAHFQREISDE 239

Query: 285 GAIFVGSPEVVAHKIIGLVEALELDRFMLHLPVGSMPHKDVLNAIKLYGKEVAPIVRKYF 344
 GA+FVGSPE VA K+I ++E L L+RFMLH+PVGSMPH+ ++ AIKLYGK V PI+ YF
 Sbjct: 240 GAMFVGSPE TVARKMIKVIEELGLNRFMLHIPVGSMPHERIMKAIKLYGKRVKPIEDYF 299

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 349

30 A DNA sequence (GBSx0380) was identified in *S.agalactiae* <SEQ ID 1131> which encodes the amino acid sequence <SEQ ID 1132>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9715> which encodes amino acid sequence <SEQ ID 9716> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1133> which encodes the amino acid sequence <SEQ ID 1134>. Analysis of this protein sequence reveals the following:

45 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0915(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 20/40 (50%), Positives = 27/40 (67%), Gaps = 3/40 (7%)

55 Query: 4 MAITHKRQDDLESMPASFAKVP---KPKKVDSDSKPEQKD 40
 MAITHK+ D+LE M A FA +P KP +V++D K K+

-442-

Sbjct: 1 MAITHKKNDELEKMLAGFASIPSFDPKPLEVNTDGKLATKE 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 350

A DNA sequence (GBSx0381) was identified in *S.agalactiae* <SEQ ID 1135> which encodes the amino acid sequence <SEQ ID 1136>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1453(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 351

A DNA sequence (GBSx0382) was identified in *S.agalactiae* <SEQ ID 1137> which encodes the amino acid sequence <SEQ ID 1138>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.15 Transmembrane 216 - 232 (210 - 240)

INTEGRAL Likelihood = -9.18 Transmembrane 15 - 31 (10 - 39)

INTEGRAL Likelihood = -9.02 Transmembrane 283 - 299 (276 - 299)

INTEGRAL Likelihood = -8.76 Transmembrane 128 - 144 (119 - 150)

INTEGRAL Likelihood = -4.62 Transmembrane 243 - 259 (237 - 265)

INTEGRAL Likelihood = -2.44 Transmembrane 65 - 81 (65 - 81)

INTEGRAL Likelihood = -2.44 Transmembrane 94 - 110 (93 - 111)

----- Final Results -----

bacterial membrane --- Certainty=0.5458(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]

Identities = 141/283 (49%), Positives = 198/283 (69%), Gaps = 3/283 (1%)

Query: 10 SVLQWFAPFISIIIEALPFVLLGTILSGIIEVFITPDIVNKFPLPKNKFRLVLFQTFVGVFV 69

S LQ +IFISI+IEA+PF+L+G ILSGII++F++ +++ + +PKN+FL VLFQ G +

Sbjct: 6 SFLQLNSIFISILIEAIPFILIGVILSGIIMFVSEEMIRIMPKNRFLAVLFGALAGVL 65

Query: 70 FPSCECGIIPINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSIRFLILRFVG 129

FP+CECGIIP I R L K VP + V F+ TAPIINPIVLF+TY AFGN + R

Sbjct: 66 FPACCEGIIPITRRLLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

Query: 130 ATIVAIALGVMLAFLVDNLIKEDAKPTHFDYSDKKWYQKIFLALAHAIDEFFDTGRYL 189

A V++ +GV+L++ DN L + +P H H + QK+ L HAIDFF G+YL

Sbjct: 126 ALAVSLIIGVILSYQFDNQLKPDEPGHHHHHGTLLQKLGGLRHAIDFFSVGKYL 184

Query: 190 VFGTLIASAMQIYLPTRVLTITGHSPITAILVMMLLAFILSLCSEADAFIGASLLSTFGI 249

-443-

+ G IA+AMQ Y+ T L IG + +++ LVMM LAF+LSLCSE DAFI +S STF +
 Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244

Query: 250 APVMAFLLLIGPMIDIKNLMMVNSFKTRFIVQFISVSSLI III 292

5

++AFL+ G M+DIKNL+MM+ +FK RF+ F+ ++ +++I+

Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMLAFAFKRFV--FLLITYIVVIV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1139> which encodes the amino acid sequence <SEQ ID 1140>. Analysis of this protein sequence reveals the following:

10

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

15

INTEGRAL	Likelihood = -9.92	Transmembrane	216 - 232 (211 - 237)
INTEGRAL	Likelihood = -9.45	Transmembrane	283 - 299 (276 - 299)
INTEGRAL	Likelihood = -8.76	Transmembrane	128 - 144 (119 - 150)
INTEGRAL	Likelihood = -7.80	Transmembrane	15 - 31 (10 - 39)
INTEGRAL	Likelihood = -5.47	Transmembrane	243 - 259 (237 - 265)
INTEGRAL	Likelihood = -2.44	Transmembrane	65 - 81 (65 - 81)
INTEGRAL	Likelihood = -2.44	Transmembrane	94 - 110 (93 - 111)

20

----- Final Results -----

bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25

The protein has homology with the following sequences in the databases:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]

Identities = 143/288 (49%), Positives = 196/288 (67%), Gaps = 1/288 (0%)

30

Query: 10 SVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPPELVQKYLKQKCLRI LFGTFVGFV 69
 S LQ +IF+SI+IEA+PF+L+G ILSG I++FV+ E++ + +PK + L +LFG G +
 Sbjct: 6 SFLQLNSIFISILIEAIPFILIGVILSGIIMFVSEEMIRIMPKNRFLAVLFGALAGVL 65

35

Query: 70 FSCCEGGIIPINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSLRFILRLVG 129
 FF+CECGIIP I R L K VP + V F+ TAPIINPIVLF+TY AFGN + R
 Sbjct: 66 FFACEGGIIPITRRLLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

40

Query: 130 AALVAITLGVMLAFIVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAIDEFFDTGRYL 189
 A V++ +GV+L++ DN L +P H H + H +L +++ L HAIDEFF G+YL
 Sbjct: 126 ALAVSLIIGVILSYQFKDNQLKPDPEGH-HHHHHGTLLQKLGGLRLHAIDEFFSVGKYL 184

45

Query: 190 VFGTLIASAMQIYVPTVRLTTIGHNPLTAILMMLMAFILSLCSEADAFIGASLLSTFGV 249
 + G IA+AMQ YV T L IG N +++ L+MM +AF+LSLCSE DAFI +S STF +
 Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244

Query: 250 APVLAFLLLIGPMVDIKNLMMVKAQKGRFIVQFISVLMIAVYCLLV 297

++AFL+ G MVDIKNL+MM+ AFK RF+ I V+++ LLV

Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMLAFAFKRFVFLITYIVVIVLAGSLV 292

An alignment of the GAS and GBS proteins is shown below:

50

Identities = 248/300 (82%), Positives = 278/300 (92%)

Query: 1 MDIFNQLPDSVLQWFAIFISIIIEALPFVLLGTILSGIIEVFITPDIVNKF LPKNKFLRV 60
 M +F+ LP SVLQWFAIF+SIIIEALPFVLLGTILSG IEVF+TP++V K+LPK K LR+
 Sbjct: 1 MSLFNSLPSSVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPPELVQKYLKQKCLRI 60

55

Query: 61 LFGTFVGFVFPSCCEGGIIPINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSI 120
 LFGTFVGFVFPSCCEGGIIPINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNS+
 Sbjct: 61 LFGTFVGFVFPSCCEGGIIPINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSL 120

60

Query: 121 RFLILRFVGVATIVAIALGVMLAFVDDNILKEDAKPTHFHDYSKKWYQKIFLALAHAI 180
 RFLILR VGA +VAI LGVMLAF+VDDNILK++A+P HFHDYS + ++I+LAL HAID
 Sbjct: 121 RFLILRLVGAALVAITLGVMLAFVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAID 180

Query: 181 EFFDTGRYLVFGTLIASAMQIYLPTRVLTIGHSPITAILVMMMLAFILSLCSEADAFIG 240

-444-

```

          EFFDTGRYLVFGLIASAMQIY+PTRVLTITIGH+P+TAIL+MML+AFILSLCSEADAFIG
Sbjct: 181 EFFDTGRYLVFGLIASAMQIYVPTRVLTITIGHNPLTAILIMMLMAFILSLCSEADAFIG 240

Query: 241 ASLLSTFGIAPVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIITIIYCLFVGVI 300
          ASLLSTFG+APV+AFLLIGPM+DIKNLMMMV +FK RFIVQFI VS L+I +YCL VGV+
Sbjct: 241 ASLLSTFGVAPVLAFLIGPMVDIKNLMMMVKAFFKGRFIVQFIGVSVMIAVYCLLVGVL 300

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 352

A DNA sequence (GBSx0383) was identified in *S.agalactiae* <SEQ ID 1141> which encodes the amino acid sequence <SEQ ID 1142>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4703 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 353

A DNA sequence (GBSx0384) was identified in *S.agalactiae* <SEQ ID 1143> which encodes the amino acid sequence <SEQ ID 1144>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL    Likelihood = -8.44    Transmembrane    45 - 61 ( 39 - 65)
INTEGRAL    Likelihood = -8.12    Transmembrane    83 - 99 ( 77 - 101)
INTEGRAL    Likelihood = -0.00    Transmembrane     2 - 18 ( 1 - 19)

----- Final Results -----
          bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8559> which encodes amino acid sequence <SEQ ID 8560> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1    Crend: 2
SRCFLG: 0
McG: Length of UR: 8
      Peak Value of UR: 2.23
      Net Charge of CR: 1
McG: Discrim Score: 0.46
GvH: Signal Score (-7.5): -3.54
      Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program count: 2 value: -8.44 threshold: 0.0
INTEGRAL    Likelihood = -8.44    Transmembrane    37 - 53 ( 31 - 57)
INTEGRAL    Likelihood = -8.12    Transmembrane    75 - 91 ( 69 - 93)
PERIPHERAL  Likelihood = 2.76      200

```

-445-

modified ALOM score: 2.19
icml HYPID: 7 CFP: 0.438

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12118 GB:Z99105 ycgQ [Bacillus subtilis]
Identities = 100/290 (34%), Positives = 159/290 (54%), Gaps = 25/290 (8%)

Query: 9 MIRFLILAGYFELSMYKLSGKLNQYINTHYTYLAYISMVLSFILAIVQLIIWVKNMKM 68
M R L+L G+ +L SG L +YIN Y YL++I++ L IL VQ +++K+ +
Sbjct: 1 MFRLLVLMGFTFFFYHLHASGNLTKYINMKYAYLSFIAIFLLAILTAVQAYLEFKSPEKS 60

Query: 69 SHLHGKIA-----KSTSP-----MILVFPVLVGLLVPTVSLDSTTVSAKGYN 110
H H + P ++ +FP++ G+ P +LDS+ V KG++
Sbjct: 61 GHHDHDCGCGHDHEHDHEQNKFYQRYLIYVVFLFPLVSGIFFPIATLDSSIVKTKGFS 120

Query: 111 FPLAAGSTGTVSQDGRVQYLPDTSYFTSSAYBKEMQKELKKYKSGTLTITTENYME 170
F A S SQ QYL+PD S Y+ +Y+K+M++ KY +++T +++++
Sbjct: 121 FK-AMESGDHYSQ---TQYLRPDASLYYAQDSYDKQMKQLFNKYSSKKEISLTDDDFLK 175

Query: 171 VMELIYLYPEQFMDROIQYTGfVY-NEPKHEGYQIFRFGIIHICIADSGVYGLLT-GNQ 228
ME IY YP +F+ R I++ GF Y ++ F+ RFGIIHICIADSGVYG+L
Sbjct: 176 GMETIYNYPGEFLGRTIEFHGFAYKGNAINKNQLFVLRFGLIHCIDSGVYGMVLEFPKD 235

Query: 229 KSYPDNTWTVVRGTIKSEYNQLLQNLPLVHLIEESRQVSKANNPYVYRVF 278
D+ W+ ++GT+ SEY Q + LPV+ + + + K ++PYVYR F
Sbjct: 236 MDIKDDEWIHIKGTLASSEYQPFKSTLPVVKVTDWNTIKKPDDEPYVYRGF 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1145> which encodes the amino acid sequence <SEQ ID 1146>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 83 - 99 (74 - 101)
INTEGRAL Likelihood = -6.21 Transmembrane 42 - 58 (39 - 62)

----- Final Results -----

bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9115> which encodes the amino acid sequence <SEQ ID 9116>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 54

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 75 - 91 (66 - 93)
INTEGRAL Likelihood = -6.21 Transmembrane 34 - 50 (31 - 54)
PERIPHERAL Likelihood = 2.76

----- Final Results -----

bacterial membrane --- Certainty= 0.433(Affirmative) < succ>
bacterial outside --- Certainty= 0.000(Not Clear) < succ>
bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/279 (74%), Positives = 244/279 (86%), Gaps = 1/279 (0%)

-446-

Query: 1 MFICGNNIMIRFLILAGYFELSMYKLKSGKLNQYINHTYTLAYISMVLSFILAIIVQLII 60
 +F CGG +MIRFLILAGYFEL+MYL+LSGKL+QYIN Y+YLAYISM+LSFILA+VQL
 Sbjct: 1 LFTCGGALMIRFLILAGYFELTMYLQLSGKLDQYINVRYSLAYISMILSFILALVQLYT 60

5 Query: 61 WVKNMKMHSHLHGKIAKSTSPMILVFPVLVGLLVPTVSLDSTTVSAKGYNFPLAAGSTGT 120
 W+KN+K+HSHL GKIA+ TSP ILVFPVL+GLLVPTV+LDSTTVSAKGY FPLAAG++ T
 Sbjct: 61 WMKNIKVHSHLTGKIARLTSPFILVFPVLIGLLVPTVTLDSSTTVSAKGYTFPLAAGASKT 120

10 Query: 121 -VSQDGRVQYLKPDSTSTYFTSSAYEKEMQKELKKYKSGTTLTITTENYMEVMELIYLYP 179
 VS DGT +QYLKPDTS YFT SAY+KEM++EL KYKG +TITTENYMEVMELIYLYP
 Sbjct: 121 GVSDDGTTIQYLKPDTSLYFTKSAYQKEMRQELHKYKGGKPVTTITTENYMEVMELIYLYP 180

15 Query: 180 EQFMDRQIQYTGfVYNEPKHEGYQFIFRFGIIHCIADSGVYGLLTGNGKSYPDNTWVTV 239
 ++F+DR IQYTGfVYNEP H+ YQF+FRFGIIHCIADSGVYGLLTGNGQ SYP+NTW+TV
 Sbjct: 181 DEFLDRDIQYTGfVYNEPGHDNYQFLFRFGIIHCIADSGVYGLLTGNGQTSYPNNTWLTW 240

20 Query: 240 RGTIKSEYNQLLQQLPVLHIEESRQVSKANNPYVYRVF 278
 +G + EY++ L+Q+LPVL + E Q + NNPYVYRVF
 Sbjct: 241 KGRLLHMEYDKNLBQHLPLVQLAEVHQTKBENNPYVYRVF 279

SEQ ID 8560 (GBS235d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 14 & 15; MW 48.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 17 & 18; MW 23.4kDa), in Figure 150 (lane 15; MW 23kDa) and in Figure 182 (lane 5; MW 23kDa).

25 GBS235d-His was purified as shown in Figure 235, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 354

30 A DNA sequence (GBSx0385) was identified in *S.agalactiae* <SEQ ID 1147> which encodes the amino acid sequence <SEQ ID 1148>. This protein is predicted to be signal recognition particle (ftsY). Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06205 GB:AP001515 signal recognition particle (docking
 protein) [Bacillus halodurans]
 Identities = 175/304 (57%), Positives = 227/304 (74%)

45 Query: 233 EKYNRSLLKTRTGFSARLNAFLSNFRVDEEFFEELEMLILSDVGVNVAQTLEDLRYE 292
 EK+ L+KTR F+ ++N + +R VDE+FFEELEE+LI +DVGVL L E+L+ E
 Sbjct: 20 EKFKAGLEKTRDSFAGKMNDLVYKYSVDEDFEELEEILIGADVGVTTVMIDLVEELKDE 79

50 Query: 293 AKLENAKKSEDLRVIVEKLVETIEYKDGTYNEAINFQGLTVMLFVGNGVGKTTSIGKL 352
 + +N K S+D++ +I EKL E+ EK+G E GL+V+L VGVNGVGKTTSIGKL
 Sbjct: 80 VRRQNIKDSKDIQPIISEKLAELLEKEGGETEVLNLPAGLSVILVVGNGVGKTTSIGKL 139

55 Query: 353 AHQYKSQGGKVMVAADTFRAGAVQLVEWGRRVDVPVVTGEEKADPASVVFDMGEKAVA 412
 AH YK QGKKV+L A DTFRAGA+ QL WG R V V+ E +DPA+V+FD ++ A +
 Sbjct: 140 AHMYKQQGKKVILAAGDTFRAGAIEQLEVWGERAGVDVIKQSEGSPPAAVMFDAIQAAS 199

Query: 413 QGVDVLLIDTAGRIQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNALSQAKE 472

-447-

+ D+L+ DTAGRLQNK NLM BLEK+ R+I R +P APHE L+ALDA+TGQNA+SQAK
 Sbjct: 200 READILICDTAGRLQNKVNLMELEKVKRVISREIPGAPHEVLIALDATTGQNAMSQAKT 259
 Query: 473 FSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKIDDIGEFNSDFMRGLL 532
 5 F + T +TG+ILTK+DGTAKGG+VLAI R ELDIPVKF+G GEKIDD+ F+SE F+ GL
 Sbjct: 260 FKETTDVDTGIIITKLDGTAKGGIVLAIRHELDIPVKFVGLGEKIDDLQPFDFSEQFVYGLF 319
 Query: 533 EGIL 536
 + ++
 10 Sbjct: 320 KDMV 323

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1149> which encodes the amino acid sequence <SEQ ID 1150>. Analysis of this protein sequence reveals the following:

Possible site: 60
 15 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4384(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 339/549 (61%), Positives = 404/549 (72%), Gaps = 46/549 (8%)

25 Query: 1 MGLFDRLFGHKKKDKEPEIEASESVLEDEDSVIDKEEGSNFSKESTLNRTSEVPVAEDD 60
 MGLFDRLFG K+ K E+ E+++ E KEE S + E ++ + +
 Sbjct: 1 MGLFDRLFGKKETPKVAEEKLEENLLTE----TTQKEELSEKANEQ-----DKIEAVQQE 51
 Query: 61 SFLELERDTALSHQPVTSIHPLESEDTDEIPVKEDDSFLELEDRAKTKVADTSEVEN 120
 30 ++ + A S + P + ++ L E+T D + DT+E
 Sbjct: 52 ---DVSSEGAGSVENGPEAASVNALVEEETG-----DNSNHPSEDITNEF-- 92
 Query: 121 VVPDSTTSLSDNVSASEASFSQSDQFSETPLQEEMS--SGKTEVQTESEDT 178
 D T L VS S+++ S+ + L D +QF Q + S S E S++
 35 Sbjct: 93 -AADKTDLK--VSELSQSTASEPKDLVDQPVVEQFPTKQAQADASNDSEAEVADTSKEQ 149
 Query: 179 SAADAFADYYAKRKAIEKEISSNSLST-----DESEFSEAQEVLSQSQA--DTIK 227
 S++ + DYY ++ A+EK + + +T E++ S + E SQ++A DTI
 40 Sbjct: 150 SSSQQVMEDYYRRKAALKEKSLQEKAATVPMPEEVPQENQASTSAEA-SQNKATHDTIP 208
 Query: 228 AESQEEKYNRSLKKTTRTGFSARLNAFLSNFRVDEEFFEELMLILSDVGVNVATQLTE 287
 E+ +EKY RSLKKTTRTGFSARLN+F +NFRVDEEFFE+LEMLILSDVGV+VAT LTE
 Sbjct: 209 -ETDQEKYKRSLLKKTTRTGFSARLNSFFANFRVDEEFFEDLEMLILSDVGVHVTTLTE 267
 Query: 288 DLRYEAKLENAKKSEDLKRVIVEKLVIEYKDG IYNEAINFQGLTVMLFVG VNGVGT 347
 +LRYEAKLENAKK + LKRVIVEKLV+IYEKDG YNEAIN+Q+GLTVMLFVG VNGVGT
 45 Sbjct: 268 ELRYEAKLENAKKPDALKRVIVEKLVDIYEKDG IYNEAINYQDGLTVMLFVG VNGVGT 327
 Query: 348 SIGKLAHQYKQKQKVMVAADTFRAGAVQLVWEGRRVDVPVVTGEEKADPASVVFDDGM 407
 SIGKLA++YK +GKKVMLVAADTFRAGAVQLVWEGRRVDVPV+TG EKADPASVVFDDGM
 50 Sbjct: 328 SIGKLAYRYKQEGKKVMLVAADTFRAGAVQLVWEGRRVDVPVITGPEKADPASVVFDDGM 387
 Query: 408 EKAVAGQGVLDLLIDTAGRLQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNAL 467
 EKAVA+GVD+LLIDTAGRLQNKENLMAELEK+GRIIKRV+PDAPHETLLALDASTGQNAL
 55 Sbjct: 388 EKAVAGVDILLIDTAGRLQNKENLMAELEKMGRIIKRVLPDAPHETLLALDASTGQNAL 447
 Query: 468 SOAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKIDDIGEFNSDF 527
 SOAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEK+DDIGEF+SEDF
 60 Sbjct: 448 SOAKEFSKITPLTGLILTKIDGTAKGGVVLAIQELDIPVKFIGFGEKVDDIGEFHSEDF 507
 Query: 528 MRGLLEGIL 536
 M+GLLEGIL
 Sbjct: 508 MKGLLEGIL 516

-448-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 355

A DNA sequence (GBSx0386) was identified in *S.agalactiae* <SEQ ID 1151> which encodes the amino acid sequence <SEQ ID 1152>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3592(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AAA62048 GB:L10328 f270 [Escherichia coli]
 Identities = 101/273 (36%), Positives = 160/273 (57%), Gaps = 10/273 (3%)

Query: 4 IKILALDLDGTLFTTDDKKVSEENKVALKAAREKGKIKVVITGRPLKAIGNLLEDLELVSD 63
 IK++A+D+DGTLD D +S K A+ AAR +G+ VV+TTGRP + N L++L +

20 Sbjct: 3 IKLIAIDMDGTLILLDPHTISPAVKNAIAAARAGVNVVLTITGRPYAGVHNYLKELHMEQP 62

Query: 64 EDYSITFNGGLVQONT-GKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYS----I 118
 DY IT+NG LVQ+ G +A+TA++ + + + +VG L T+Y+ I

25 Sbjct: 63 GDYCITYNGALVQKAADGSTVAQTALSYYDDYRXLEKLSREVGSHFHALDRTTLYTANRDI 122

Query: 119 ANKGHHSQYHLANPLLEFIEVDLDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDY 178
 + H + PL+ F E E++ + + K++ + + LDQ IA++P +K Y

30 Sbjct: 123 SYTYVHESFVATIPLV-FCEA---EKMDPNTQFLKVMIDEPAILDQAIARIPQXVKEKY 178

Query: 179 EMFKSRDIILELMPKGVHKAAGVLELLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVAMA 238
 + KS LE++ K V+K G++ L LG+ ++MA+GD+ ND++M+E+AG+GVAM

35 Sbjct: 179 TVLKSAPYFLEILDKRNVNKGTVGKSLADVLGKPEIIMAIGDQENDIAMIEYAGVGVMAD 238

Query: 239 NGIPEAKAIAKATTICNNDSEGVAAEIGKYILS 271
 N IP K +A T +N E GVA AI KY+L+

35 Sbjct: 239 NAIPSVKEVANFVT-KSNLEDGVAFAIEKYVLN 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1153> which encodes the amino acid sequence <SEQ ID 1154>. Analysis of this protein sequence reveals the following:

40 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3502(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 180/273 (65%), Positives = 218/273 (78%), Gaps = 1/273 (0%)

Query: 3 DIKILALDLDGTLFTTDDKKVSEENKVALKAAREKGKIKVVITGRPLKAIGNLLEDLELVSD 62
 +I+ILALDLDGTL+ T+K V++ NK AL AAREKG+KVVITGRPLKAIGNLLE+L+L+

55 Sbjct: 2 NIRILALDLDGTLTYNTEKIVTDANKKALAAAREKGKGVKVVITGRPLKAIGNLLEEDLLD 61

Query: 63 DEDYSITFNGGLVQONTGKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYSIANK- 121
 +DYSITFNGGLVQ+NTG++L K++++ +V I + L VGLPTDI+S G VYSI +K

60 Sbjct: 62 HDDYSITFNGGLVQRNTGEVLDKSSLSFDQVCQIQALEAVGLEPTDIISGGDVYSIPSKD 121

Query: 122 GHHSQYHLANPLLEFIEVDLDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDYEMF 181

-449-

G HSQYHLANPLL FIEV + ++PKD+ YNKIV+V D +LDQQI KI L D+E F
 Sbjct: 122 GRHSQYHLANPLLTFIEVTSVAELPKDITYNKIVTVTDPDFLDQQI IKLSPSLFEDFRAF 181

Query: 182 KSRDIIILELMPKGVHKAVGLELLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVMANGI 241
 KSRDII E+MPKG+ KA GL LL +HLGLD+ VMAMGDEAND +MLEWAGLGVMANG+
 Sbjct: 182 KSRDIIIFEIMPKGIDKAFGLNLLCQHLGLDARHVMAMGDEANDFAMLEWAGLGVMANGV 241

Query: 242 PEAKAIAKATTICNDESGVAEAIGKYILSEEN 274
 AKA A A T NDESGVAEA+ +IL EE+
 Sbjct: 242 SGAKADADAVTTLTNDESGVAEAVKTFILEEES 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 356

- 15 A DNA sequence (GBSx0387) was identified in *S.agalactiae* <SEQ ID 1155> which encodes the amino acid sequence <SEQ ID 1156>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4648(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
 intergenic region. [Escherichia coli]
 Identities = 91/264 (34%), Positives = 146/264 (54%), Gaps = 4/264 (1%)

30 Query: 2 IKLVATDMDGTFLDENGTYDKKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFQDM 61
 IKL+A DMDGTFL + TY+++R ++ K QGI F ASG L F + +++
 Sbjct: 4 IKLIAVDMDGTFLSDQKTYNRERFMAQYQOMKAQGI RFVVASGNQYYQLISFFPEIANEI 63

35 Query: 62 AFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMNENYVLSGKDGAYILSDAN 121
 AF+AEENG V + + LS++ + +++HL P + E + GK+ AY L +
 Sbjct: 64 AFVAENGWVWVSEKDVFNGLSKDAFATVVEHLLTRPEV---EIIACGKNSAYTLKKYD 120

40 Query: 122 PDYIEFITHYYDNLQKVSHFEDVDDIIFKVTANFTEETVRQAEWVNQAI-PYATAVTTG 180
 YY L+ V +++++DI FK N ++E + Q ++ +++AI +V TG
 Sbjct: 121 DAMKTVAEMYHRLEYVDNFDNLBIDIFFKFGNLNDELIPQVQKALHEAIGDIMVSVHTG 180

45 Query: 181 FKSIDIILSSVKNRNGLEHLCEQYGIRAEVLSFGDNINDLEMLEWSGKAIATENARPEV 240
 SID+I+ V+K NGL L + +GI EV+ FGD ND+EML +G + A ENA V
 Sbjct: 181 NGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSAV 240

Query: 241 KEIADCIIGHNNQAVMAYLESMV 264
 A G +N + V+ ++ ++
 Sbjct: 241 VAAAKYRAGSNNREGVLDVIDKVL 264

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1157> which encodes the amino acid sequence <SEQ ID 1158>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 60 An alignment of the GAS and GBS proteins is shown below:

-450-

Identities = 138/265 (52%), Positives = 193/265 (72%), Gaps = 1/265 (0%)

Query: 1 MIKLVATDMDGTFLDENGTYDKKRLANVLKKFKEQGIVFTASGRSLLSLEQLFADFRDQ 60
 MIKL+ATDMDGTFL E+GTY++++LA +L K E+GI+F +SGRSLL+++QLF F DQ
 5 Sbjct: 1 MIKLIATDMDGTFLAEDGTYNQEQLAALLPKLAEGILFAVSSGRSLLAIDQLFEPFLDQ 60

Query: 61 MAFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMNENYVLSGKDGAYILSDA 120
 +A IAENGs + + +++EQY ++ + +P+ V SG+ AYIL A
 10 Sbjct: 61 IAVIAENGSVVQYRGEILFADMMTKEQYTEVAKKILANPHYVETGMVFSGQKAAYILKGA 120

Query: 121 NPDYIEFITHYYDNLQKVSHFEDVD-DIIFKVTANFTEETVRQAEWVNQAIPYATAVTT 179
 + +YI+ HYY N++ ++ FED++ D IFKV+ NFT TV + +W+NQA+PYATAVTT
 Sbjct: 121 SEEYIQTKHYHYNVQVINGFEDMENDAIFKVSTNFTGHTVLEGSWLNQALPYATAVTT 180

Query: 180 GFKSIDIIILSSVNKRNGLEHLCEQYGIKRAEVLSPFGDNINDLEMLEWSGKAIATENARPE 239
 GF SIDIIL VNK G+EHLC+ GI+ E ++FGDN ND +MLE++G+AIATENARPE
 15 Sbjct: 181 GFDSIDIILKEVNKGFGMEHLCQALGIKKAETIAFGDNFNDYQMLEFAGRAIATENARPE 240

Query: 240 VKEIADCIIGHNNQAVMAYLESMV 264
 +K I+D +IGH N+ AV+ YL+ +V
 20 Sbjct: 241 IKVISDQVIGHCNDGAVLTYLKGVLV 265

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 357

A DNA sequence (GBSx0388) was identified in *S.agalactiae* <SEQ ID 1159> which encodes the amino acid sequence <SEQ ID 1160>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2428(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 358

A DNA sequence (GBSx0389) was identified in *S.agalactiae* <SEQ ID 1161> which encodes the amino acid sequence <SEQ ID 1162>. This protein is predicted to be p115 protein (smc). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 50

A related GBS nucleic acid sequence <SEQ ID 9713> which encodes amino acid sequence <SEQ ID 9714> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg
[Bacillus subtilis]

Identities = 458/1193 (38%), Positives = 728/1193 (60%), Gaps = 27/1193 (2%)

5 Query: 1 MFLKEIEMQGFKSFADKTKVEFDQGVTA VVGPNVSGKSNITESLRWALGESSAKSLRGGK 60
MFLK +++ GFKSFA++ V+F +GVTAVVGPNGSGKSNIT+++RW LGE SA+SLRGGK
Sbjct: 1 MFLKRLDVIGFKSFAERISVDVFKGVTA VVGPNVSGKSNITDAIRWVLGEQSARSLRGGK 60

10 Query: 61 MPDVIFAGTFENRKPLNYA QSVTL DNDSDHFIENIAD EVRVERRIFRNGDSEYLIDGRKVR 120
M D+IFAG+++RK LN A+V++TLDN DHF+ EV V RR++R+G+SE+LI+ + R
Sbjct: 61 MEDIIFAGSDSRKRLNLA E VTLTLDNDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120

15 Query: 121 LRDIHDLFMDITGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKETQSK 180
L+DI DLFMD+GLG+++FSIISQG+VE I +SK E+RR+IFEEAAGVLKYKTRKK+ ++K
Sbjct: 121 LKDIIDL FMD SGLGKEAFSII SQGKVEEILSSKAEDRRSIFEEAAGVLKYKTRKKKAENK 180

20 Query: 181 LEQTQGNLDRLEDIYELDMQVQPLEKQASTAKRFLVLDEERQGLHLSILIEDILQHQS D 240
L +TQ NL+R+EDI++EL+ QV+PL+ QASTAK +L +E + + +++ DI +
Sbjct: 181 LFETQDNLNRVEDILHELEQGVEPLKI QASTAKDYLEKKKELEHVEIALTAYDIEKLHGK 240

25 Query: 241 LTTVEEKLTVRKELATYYQQRQSLDENQSLKQKRHHLSEEIEAKQLALLDVTKLKSDL 300
+T++EK+ ++E + E + + + K L E + Q LL ++ L
Sbjct: 241 WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDESVELQVLLVLTSEELEKL 300

30 Query: 301 ERQIDLIRLESNQKAEKKEEAGQRLAELEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
E + +++ + +E+ + + + + K ++++++ TL ++ Q+R+++
Sbjct: 301 EGRKEVLKERKKNVQVQEQLEEAIVQFQKQETVLKEELSKQEA V FETLQAEVQQLRAQV 360

35 Query: 361 VSTESSLERFSTNPDQIIEKLREDFVTLMQEEADTSNALTALLADIENQKQASQAKSQEI 420
+ +L + N ++ IE+L+ D+ L+ +A N L LL D +Q + + +
Sbjct: 361 KEKQALSLHNENVEEKIEQLKSDYFELLNSQASIRNEL-QLDDQMSQSAVTLQRLADN 419

40 Query: 421 QEVSKNLEVLKSNKVALE-RFEAAKKNVRQLLSHYQDLGQTLQNLGEYKNNQSSILFDH 479
E S K A E F ++ + + Y+D+ + + +Y+ +S L+
Sbjct: 420 NEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEQKKRQYKKNESALYQA 479

45 Query: 480 LDEIKSQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALE 539
++ +++ LE++ + S FY GVK VL+AK++LGGI GAV E +S ++ Y+TA+E
Sbjct: 480 YQYVQQA RSKDMLETMQGFSGFYQGVKEVLKAKERLGGIRGAVLELISTEQKYETAIE 539

50 Query: 540 IALGSSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHLYSKLQSSQGF 599
IALG S+QH++ +DE +A+++I +LK+N GRATFLPL+ I+ R+L F
Sbjct: 540 IALGASAHVVTDEQSAKRAIQYLKQNSFGGRATFLPLSVIRDRQLQSRDAETAARHSSF 599

55 Query: 600 LGIASSELVTYDQRLSNIFKNNLGLTAIFDTVDNANVAARQLNYQVRLVTLTGTEL RP GGS 659
LG+ASELVT+D ++ +N LG I + + AN A+ L ++ R+VTL+G + PGGS
Sbjct: 600 LGVASELVTDFPAYRSVIQNLGLTVLITEDLKGANELAKLLGHRYRIVTLEGDVVNPGGS 659

60 Query: 660 YSGGANRQNTVFI--KPELDNLKKELKQAQSKQLIQEKEVATLLEQLKEKETLAQLKN 717
+GGA ++ N + EL+++ K L + + K + E+EV TL +++ ++ LA L+
Sbjct: 660 MTGGA VKKKNNSLLGRSRELEDTVTKRLAEMEETALLEQEVKTLKHSTIQDMEKKLADLRE 719

65 Query: 718 DGEQARLEBQRADEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENE--KNRLEKELEQ 775
GE RL++Q + +L ++N AL ++ E + K +LE+EL
Sbjct: 720 TGEGLRLKQDVKGLYELQVAEKNINTHELYDQEKSALESSEDEERKVRKRKLEBELSA 779

70 Query: 776 FAIKKEELTTSIAQIKEDKDSIQBKVNNTLLSEALBERDLLNEQKFERANCTRL--- 832
+ K ++L I ++ + K + +L+ L+E ++ K E N RL
Sbjct: 780 VSEKMKQLEEDIDRLTKQKQTSSTKESLSNELTELKIAAAKKEQACKGEEDNLARLKKE 839

75 Query: 833 ----EITLSEIKRDISNLQTLSSHQDSQLDKEELPRIEQQLQVNNRRENDEEKLVS LRF 888
E+ L E K D+S L + +S S E++L + + ND+ K + L
Sbjct: 840 LTETELALKEAKEDLSFLTSEMSSSTSG-----EEKLEEA AKHKLNDKTKTIELIA 890

80 Query: 889 ELEDCEAALDDLAASLAKEGQKNESLIRQQAQL---ESQCEQLSQQLMIFSRQLSEDYQ 944
D L + +E ++ + L +Q+ L E + ++ +L + L E+Y

-452-

5 Sbjct: 891 LRRDQRIKLQHLDTYERELKEMKRLYKQKTLLKDEEVKLGMEVELDNLLQYLREEYS 950
 Query: 945 MTLDEAKVKANVLEDILMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLITFLNTQRDD 1004
 ++ + AK K + D AR+++K ++ I+ LG VN+ +I +FE V+ER FL+ Q++D
 10 Sbjct: 951 LSFEGAKEKYQLETDPEEARKRKVLIKLAIEELGTVNLSIDEFERVNERYKFLSEQKED 1010
 Query: 1005 LVHAKNLLLETITDMDEVKTRFKSTFEAIRHSFKETTFVQMFGGGSADLILTE-GDLLSA 1063
 L AKN L + I +MD+E+ RF TF IR F + F +FGGG A+L LT+ DLL +
 15 Sbjct: 1011 LITEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGRAELRLTDPNDLLHS 1070
 Query: 1064 GVDISVQPPGKKIQSLNLSGGGEKALSALALFAIRVKTIPIFVILDEVEAALDEANVKR 1123
 GV+I QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV R
 Sbjct: 1071 GVEIIAQPPGKKLQNLNLSGGERALTAIALLFSSILKVRPVPFCVLDEVEAALDEANVFR 1130
 20 Query: 1124 FGDYLNRFDKSSQFIVVTHRKGTMSSAADSIYVTMQESGVSKIVSVKLKEAQE 1176
 F YL ++ +QFIV+THRKGTM AD +YGVVTMQESGVSK++SVKL+E +E
 Sbjct: 1131 FAQYLKKYSSDTQFIVITHRKGTMEEADVLYGVVTMQESGVSKIVSVKLEETKE 1183

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1163> which encodes the amino acid
 20 sequence <SEQ ID 1164>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)
 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg
 [Bacillus subtilis]
 35 Identities = 441/1192 (36%), Positives = 729/1192 (60%), Gaps = 25/1192 (2%)
 Query: 1 MFLKEIELEGFKSFAADTKIEFDKGVTA VVGPNNGSGKSNITESLRWALGESSAKNLRGGK 60
 MFLK +++ GFKSFA++ ++F KGVTA VVGPNNGSGKSNIT+++RW LGE SA++LRGGK
 Sbjct: 1 MFLKRDLVIGFKSFAERISVDFVKGVTAVVGNNGSGKSNITDAIRWVLGEQSARSRLRGK 60
 40 Query: 61 MPDVIFAGTQNRNPLNYAKVAVLDNSDHFITAKKEIRVERHIYRNGSDSYLDGRKVR 120
 M D+IFAG+ +R LN A+V + LDN DHF+ E+ V R +YR+G+S++LI+ + R
 Sbjct: 61 MEDIIFAGSDSRKRLNLAEVTLTLDNDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120
 45 Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVBEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180
 L+DI DLFMD+GLG+++FSIISQG+VEEI +SK E+RR+IFEEAAGVLKYKTRKK+ + K
 Sbjct: 121 LKDIIDLFDMSGLGKEAFSIIISQGVBEILSKAEDRRSIFEEAAGVLKYKTRKKKAENK 180
 Query: 181 LNQTQDNLDRLEDIIYELDTQLAPLEKQAKVAKQFLELDANRQQLQLDILVKDIDIAQER 240
 L +TQDNL+R+EDI++EL+ Q+ PL+ QA +AK +LE + +++ + DI+ +
 50 Sbjct: 181 LFETQDNLNRVEDILHELEGQVEPLKIQASTIAKDYLEKKKELEHVEIALTAYDIEKLHGK 240
 Query: 241 QTKDTEALALQQDLASYYAKRQSMEEYQKFKQKQVLSQESDQTQTTLLELTCLIADL 300
 + E + +++ + + E + + K Q L + ++ Q LL ++ + L
 55 Sbjct: 241 WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDESVELQVLLVTSBELEKL 300
 Query: 301 EKQIELVKLESGQEAEEKAEAKHLEQLQEQLDGFQAEKQCTEQLLH-----IDQQQL 353
 E + E++K E+K A ++ EQL+E + FQ +E E+L + ++
 Sbjct: 301 EGRKEVLK-----ERKKNVQVQEQLEEAIVQFQKQETVLKEELSKQEAVFETLQAEV 353
 60 Query: 354 CDVKKQLNELSNALERFSSDPDQLMETLREEFVLLMQKEAALSQLTALKAHLDKEKQAR 413
 ++ Q+ E AL + + ++ +E L+ ++ L+ +A++ N+L L + +
 Sbjct: 354 KQLRAQVKEKQQALSLENENVEEKIEQLKSDYFELNLSQASIRNELQLLDDQMSQSAVTL 413
 65 Query: 414 QHKAQEYQLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQ 473
 Q A + + + ++ + + +++++ + Y++ + ++ +R Y+ N+

-453-

Sbjct: 414 QRLADNNEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEQKKRQYЕКNE 473

Query: 474 ERLFDLLDQKKGEARKASLESIQKSHSQFYAGVRAVLQSQKLGIGAVSEHLSFDS 533
L+ + ++K LE++Q S FY GV+ VL++++LGGI GAV E +S +

5 Sbjct: 474 SALYQAYQYVQOARSKKDMLETMQGFSGFYQGVKEVLKAKERLGGIRGAVLELISTEQK 533

Query: 534 YQTALEVALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQL 593
Y+TA+E+ALGA++QH++ DE +A++AI YLK+N GRATFLPL+ I+ R L

10 Sbjct: 534 YETAIEIALGASAQHVVTDEQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETA 593

Query: 594 ATCEGYLGTAEGLIRYDDSLSAIIQNLLSSTAI FETIDQANIAARLLGYKVRIVTLDGTE 653
A +LG A L+ +D + ++IQNLL + I E + AN A+LLG++ RIVTL+G

Sbjct: 594 ARHSSFLGVASELVTFDPAYRSVIQNLLGTVLITEDLKGANELAKLLGHRYRIVTLEGDV 653

15 Query: 654 LRPGGSFSGGANRQSNSTTFI--KPELEQISEELTRIVEQLKITEKEVAALQSDLIKKEE 711
+ PGGS +GGA ++ N + + ELE +++ L + E+ + E+EV L+ + +++

Sbjct: 654 VNPGGSMITGGAVKKNNLSLGRSRELEDTVTKRLAEMEETALLEQEVKTLKHSIQDMEKK 713

Query: 712 LTQKLQAGDQARLAEQ--RAQMAYQQLQEQKQEDSKALLAALDQSQTTSHDESLLAEQARI 769
L L+ G+ RL +Q + Q+ Q+ EK ++ L ++S + SDE + ++

20 Sbjct: 714 LADLRETGEGRLKQQDVKGQLYELQVAEKNINTHLELYDQEKSALESSEDERKVRKRKL 773

Query: 770 EEALTAIAKKKNALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDILLNEKKFEQANQ 829
EE L+A+++K L DID + + K +++ L++ ++ K E+ N

25 Sbjct: 774 EEELSAVSEKMKQLEEDIDRLTKQKTQSSTKESLSNELTELKIAAAKKEQACKGEEDNL 833

Query: 830 SRLRTQLKQCQONILKLESILNNVNSQDSIQRLPQWQKQLODATEHKSGAQKRLVQLRFE 889
+RL+ +L + + + + + L+ S+ S +++L++A +HK + + ++L

30 Sbjct: 834 ARLKKELTETELALKEAKEDLSFLTSEMSS--TSGEKELEAAKHLNDKTKTIELIAL 891

Query: 890 IEDVEARLEETAEKITKESEKNDTFIRQTKL---ETHLEQVANRLRAYAKSLSEDFQM 945
D +L+ + +E ++ +++T L E L ++ L + L E++ +

35 Sbjct: 892 RRDQRILQHGLDITYERELKEMKRLYKQKTTLLKDEEVKLGRMEVELDNLQYLRREYSL 951

Query: 946 TLADAKEVTNSIDHLESAKEKLHHLQKTIRALGPINSDAINQYEEVHERLFTLSQKTDL 1005
+ AKE E A++++ ++ I LG +N +I+++E V+ER FL+ QK DL

40 Sbjct: 952 SFEGAKEYQLETDPEARKRVKLIKLAIEELGTVNLGSDIEFERVNERYKFLSEQKEDL 1011

Query: 1006 TKAKNLLLETINSMDSEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTE-TDILSAG 1064
T+AKN L + I MD E+ RF TF I+ F + F +FGGG A+L LT+ DLL +G

45 Sbjct: 1012 TEAKNTLFQVIEEMDEMTKRFNDTFVQIRSHFDQVFRSLFGGGRAELRLTDPNDLLHSG 1071

Query: 1065 IEISVQPPGKKIQSLNLMGGEKALSALALLFAIIRVKTIPIFVILDEVEAALDEANVKRF 1124
+EI QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV RF

50 Sbjct: 1072 VEIIAQPPGKKLQNLNLLSGGERALTALLESILKVRPVPFCVLDEVEAALDEANVRF 1131

Query: 1125 GDFLNRFDKDSQFIVVTHRKGTMMAADSIYGITMQESGVSKIVSVKLEAQE 1176
+L ++ D+QFIV+THRKGTM AD +YG+TMQESGVSK++SVKL+E +E

Sbjct: 1132 AQYLKKYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKIVSVKLEETKE 1183

An alignment of the GAS and GBS proteins is shown below:

Identities = 732/1179 (62%), Positives = 911/1179 (77%)

55 Query: 1 MFLKEIEMQGFKSFADKTKVEFDQGVTA VVGPNNGSGKSNITESLRWALGESSAKSLRGGK 60
MFLKEIE++GFKSFADKTK+EFD+GVTAVVGPNNGSGKSNITESLRWALGESSAK+LRGGK

Sbjct: 1 MFLKEIELEGFKSFADKTKIEFDKGVTA VVGPNNGSGKSNITESLRWALGESSAKNLRGGK 60

Query: 61 MPDVIFAGTENRKPLNYAQVSVTLNDSHFIENTADEVRVERIFRNGDSEYLDIGRKVR 120
MPDVIFAGT+NR PLNYA+V+V LDNSDHF+ E+RVER I+RNGDS+YLDIGRKVR

60 Sbjct: 61 MPDVIFAGTQNRNPLNYAKVAVVLDNSDHFITAKKEIRVERHIYRNGDSYLDIGRKVR 120

Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKETQSK 180
LRDIHDLFMDTGLGRDSFSIISQGRVE IFNSKPEERRAIFEEAAGVLKYKTRKKETQ K

65 Sbjct: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180

Query: 181 LEQTQGNLDRLEDIIYELDMQVQPLEKQASIAKRFLVLDEERQGLHLSILIEDILQHQS 240
L QTQ NLDRLLEDIIYELD Q+ PLEKQA +AK+FL LD R+ L L IL++DI Q

- Sbjct: 181 LNQTQDNLDRLIEDIIYELDTQLAPLEKQAKVARQFLELDANRKQLQLDILVKDIDIAQER 240
- Query: 241 LTTVEEKLTVRKELATYYQQRQSLDENQSLKQKRHHLSSEIEAKQLALLDVTKLKSDL 300
T E L ++++LA+YY +RQS+E++ Q KQK+ LS+E + Q LL++TKL +DL
- 5 Sbjct: 241 QTKDTEALALQODLASYYAKRQSMEDYQKFKQKQVLSQESDQTQTTLELTCLIADL 300
- Query: 301 ERQIDLIRLESNQAEKKEEAGQRLAELEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
E+QI+L++LES Q+AEKK EA + L +L+ + + Q +L + +++ ++ ++
- 10 Sbjct: 301 EKQIELVKLESGQRAEKKAEEKHLEQLQEQLDGFQAEKQCTEQLLHIDQQLCDVKQQL 360
- Query: 361 VSTESSLERFSTNPDQIIEKLREDFVTLMQEEADTSNALTALLADIENKQASQAKSQEI 420
++LERFS++PDQ++E LRE+FW LMQ+EA SN LTAL A ++ +KQA Q K+QE
- Sbjct: 361 NELSNALERFSSDPQLMETLREEFVLLMQKEAALSNQLTALKAHLDKEKQARQHKAEY 420
- 15 Query: 421 QEVSKNLEVLKSNKVALERFEAAKKNVRQLLSHYQDLGQTLQNLGEYKNQCSILFDHL 480
Q + L+ L ++ A ++A K+ V LL +YQ+ + +Q LE +Y+ Q LFD L
- Sbjct: 421 QLLVTKLQDLNDESQKQAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQERLFDLL 480
- Query: 481 DEIKSQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALAI 540
D+ K K+AR +SLESI K+HS FYAGV++VLQ++ +LGGIIGAVSEHLSFD YQTALE+
- 20 Sbjct: 481 DQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKKGIIIGAVSEHLSFSDSYQTALAI 540
- Query: 541 ALGSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHLYLSKLQSSQGF 600
ALG +SQHIIV DE+AAKR+IA+LKKNRQGRATFLPLTTIK R L++HY +L + +G+L
- 25 Sbjct: 541 ALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQLATCEGYL 600
- Query: 601 GIASLVTYDQRLSNIFKNNLGLTAIFDVTDNANVAARQLNYQVRLVTLTGTELPRPGGSY 660
G A L+ YD LS I +N L TAIF+T+D AN+AAAR L Y+VR+VTLTGTELPRPGGS+
- 30 Sbjct: 601 GTAESLIRYDDLSAIIQNLSSTAIFETIDQANIAARLLGYKVRIVTLTGTELPRPGGSF 660
- Query: 661 SGGANRQNTVFIPKPELDNLKELKQAQSKQLIQEKEVATLLEQLKEKQETLAQLKNDGE 720
SGGANRQ+NT FIKPEL+ + +EL + + I EKEVA L L K+E L QLK G+
- Sbjct: 661 SGGANRQSNNTFIKPELEQISEELTRLVEQLKITEKEVAALQSDLIKKEELTQLKLAGD 720
- 35 Query: 721 QARLEEQRADIEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENEKNRLEKELEQFAIKK 780
QARL EQRA + YQQL EK D L L S + E+ R+E+ L A KK
- Sbjct: 721 QARLAEQRAQMAQQLEKQEDSKALLAALDQSQTHSDESLAEQARIEEALTAIAKKK 780
- Query: 781 EELTTSIAQIKEDKDSIQEKVNNLTLLSEAELEERDILLNEQKFERANCTRLEITLSEIK 840
LT I IKR+KD I++K N+ LS+A+L+ERDLNE+KFE+AN +RL L + +
- 40 Sbjct: 781 NALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDILLNEKKFEQANQSRRLTQLKQCQ 840
- Query: 841 RDISNLQTLTLLSHQDSQLDKBELPRIEKQLLQVNNRRENDEEKLVSRLFELEDCEALDDL 900
++I L+++L++ SQ + LP+ +KQL + +++LV LRFE+ED EA L++
- 45 Sbjct: 841 QNIIKLESILNNVSQDSIQRLPQWQQLQDATEHKSGAQKRLVQLRFEIEDYEARELET 900
- Query: 901 AASLAKEGQKNESLIRQQAQLESQCEQLSQQLMIFSRQLSQEDYQMTLDEAKVKANVLEDI 960
A + KE +KN++ IR+Q +LE+ EQ++ +L +++ LSED+QMTL +AK N ++ +
- 50 Sbjct: 901 AEKITKESEKNDTFIRRTKLETHLEQVANRLRAYAKSLSEDFQMTLADAKEVTNSIDHL 960
- Query: 961 LMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFLNTQRDDLVHAKNLLLETTIDMD 1020
A+E+L LQ I+ALGP+N DAI Q+EEVHERLTFL +Q+ DL AKNLLLETTI MD
- Sbjct: 961 ESAKEKLHHLQKTIRALGPINSDAINQYEEVHERLTFLTSQKTDLTAKNLLLETTINSMD 1020
- 55 Query: 1021 DEVKTRFKSTFEAIRHSFKETFFVQMFGGGSADLILTEGDLISAGVDISVQPPGKKIQSLN 1080
EVK RFK TFEAI+ SFKETF QMFGGGSADL+LTE DLLSAG++ISVQPPGKKIQSLN
- Sbjct: 1021 SEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTETDILLSAGIEISVQPPGKKIQSLN 1080
- Query: 1081 LMSGGEKALSALALLFAIRVKTIPFVILDEVEAALDEANVKRFGDYLNRFDKSSQFIVV 1140
LMSGGEKALSALALLFAIRVKTIPFVILDEVEAALDEANVKRFGD+LNRFDK SQFIVV
- 60 Sbjct: 1081 LMSGGEKALSALALLFAIRVKTIPFVILDEVEAALDEANVKRFGDFLNRFDKSSQFIVV 1140
- Query: 1141 THRKGTMASADSIYGVIMQESGVSKIIVSVKLKEAQEMTN 1179
THRKGTM+AADSIYG+TMQESGVSKIIVSVKLKEAQEMTN
- 65 Sbjct: 1141 THRKGTMASADSIYGITMQESGVSKIIVSVKLKEAQEMTN 1179

SEQ ID 1162 (GBS199) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 2; MW 75kDa).

GBS199-GST was purified as shown in Figure 208, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 359

A DNA sequence (GBSx0390) was identified in *S. agalactiae* <SEQ ID 1165> which encodes the amino acid sequence <SEQ ID 1166>. This protein is predicted to be ribonuclease III (*rnc*). Analysis of this protein sequence reveals the following:

```

10 Possible site: 46
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
15         bacterial cytoplasm --- Certainty=0.3372 (Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9711> which encodes amino acid sequence <SEQ ID 9712> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB13466 GB:Z99112 ribonuclease III [Bacillus subtilis]
  Identities = 115/230 (50%), Positives = 154/230 (66%), Gaps = 1/230 (0%)

Query: 13  KMKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLNISHNERLEFLGDAVLQLLI 72
25      KK+++ +   E+   + F N++LL AFTH+SY NEHR      NERLEFLGDAVL+L I
Sbjct: 15  KKVEQFKKEQER-LSVHFQNEKLLYQAFTHSSYVNEHRKKPYEDNERLEFLGDAVLELTI 73

Query: 73  SQYLFTKYYPQKAEGDLSKLRSMIVREESLAGFSRLCGFDHYIKLGKGEESGGRNRDTIL 132
30      S++LF KYP +EGDL+KLR+ IV E SL +   F + LGKGEE +GGR R +L
Sbjct: 74  SRFLFAKYFAMSEGDLTKLRAAIVCEPSLVSLAHELSTFGDLVLLGKGEEMTGGRKRPALL 133

Query: 133 GDLFEAFGLGALLLDKGVVVHAFVNKVMIPHVEKGYERVKDYKTSLQELLQSHGDVKID 192
35      D+FEAF+GAL LD+G+E V +F+ + P + G + V D+K+ LQE +Q G ++
Sbjct: 134 ADVFEAFIGALYLDQGLEPVSFSLKVYVFPKINDGAFSHVMDFKSQLQEYVQRDQKGSLE 193

Query: 193 YQVTNESGPAHAKEFEVTVSVNQENLSQIGRSKKAEEQDAAKNALATLQ 242
40      Y+++NE GPAH +EFE VS+ E L G GRSKK AEQ AA+ ALA LQ
Sbjct: 194 YKISNEKGPANHREFEAIIVSLKGEPLGVNGRSKKEAEQHAAQEALAKLQ 243

```

40 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1167> which encodes the amino acid sequence <SEQ ID 1168>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1414 (Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 170/227 (74%), Positives = 192/227 (83%)

Query: 15 MKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLNISHNERLEFLGDAVLQLLISQ 74

-456-

MK+L L + I F + LL+TAFTHTSYANEHRLN+SHNERLEFLGDAVLQL+IS+
 Sbjct: 1 MKQLEELLSFSDIQFNDLTLETAFTHTSYANEHRLN+SHNERLEFLGDAVLQLIISE 60
 Query: 75 YLFTKYPQKAEGLSKLRSMIVREESLAGFSRLCGFDHYIKLGKGEKSGGRNRDITLGD 134
 5 YLF KYP+K EGD+SKLRSMIVREESLAGFSR C FD YIKLGKGEKSGGR RDTILGD
 Sbjct: 61 YLFAKYPKTEGMSKLRSMIVREESLAGFSRFSFDAVIKLGKGEKSGGRRRDTILGD 120
 Query: 135 LFEAFLLGALLLDKGEVHVHAFVNVKVMIPHVEKGTVERVKDYKTSLQELLQSHGDVKIDYQ 194
 LFEAFLLGALLLDKG++ V F+ +VMIP VEKG +ERVVDYKT LQE LQ+ GDV.IDYQ
 10 Sbjct: 121 LFEAFLLGALLLDKGIDAVRRFLKQVMI PQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQ 180
 Query: 195 VTNESGPAHAKEFEVTVSVNQENLSQIGRSKKAEEQDAAKNALATL 241
 V +E GPAHAK+FEV++ VN LS+G+G+SKK AEQDAAKNALA L
 15 Sbjct: 181 VISEKGPAAKQFEVSIVVNGAVLSKGLGSKKLAEQDAAKNALAQL 227

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 360

A DNA sequence (GBSx0391) was identified in *S.agalactiae* <SEQ ID 1169> which encodes the amino acid sequence <SEQ ID 1170>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -4.19 Transmembrane 100 - 116 (99 - 117)
 25 INTEGRAL Likelihood = -2.44 Transmembrane 81 - 97 (81 - 97)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAC12789 GB:AJ279090 hypothetical protein [Staphylococcus
 carnosus]
 Identities = 50/114 (43%), Positives = 72/114 (62%)
 35 Query: 3 KIFYISLGFISLGIGIAGIVLPVPTTPLVLLSAFCFSRSEKFDIWLRTKVYKYAAD 62
 K ++LG I GIG GIV+P++PTTP +LL+A CFSRSS+KF+ WL TK++ Y
 Sbjct: 2 KYVLTILGLIFAGIGFVGIVVPLPTTFFLLAAICFSRSSKFNRLVNTKIHDEYVES 61
 40 Query: 63 FVESRSIAPARKKSMIWQIYILMGISIYFAPLMWLKGLLIGTIVGTIVLFV 116
 F + +K ++ +YILMGISI+ +++++ LLI V T VLF V
 Sbjct: 62 FKRDKGFTLKKFKLLTSLYILMGISIFIIDNLYIRITLLIMLFVQTIVVLFV 115

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 361

A DNA sequence (GBSx0392) was identified in *S.agalactiae* <SEQ ID 1171> which encodes the amino acid sequence <SEQ ID 1172>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.1908(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-457-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1173> which encodes the amino acid sequence <SEQ ID 1174>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1610(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 225/269 (83%), Positives = 248/269 (91%)

Query: 1 MSEGFKYSILASGSTGNCFYIETPQKRLIDAGLTGKKVTSLLAEINRKPEDLDAILVT 60
 M+E GFKYSILASGSTGNCFY+ETP+KRLIDAGLTGKK+TSLAEI+RKPEDLDAIL+T
 Sbjct: 1 MNESGFKYSILASGSTGNCFYLETTPKKRLIDAGLTGKKITSLLAEIDRKPEDLDAILT 60

Query: 61 HEHSDHIKGVGV LARKYHLDIYANEQTWKVMDERNMLGKVDVSQKHVFGRGKILTFGDLD 120
 HEHSDHIKGVGV+ARKYHLDIYANE+TW++MDE NMLGK+D SQKH+F R K LTFGD+D
 Sbjct: 61 HEHSDHIKGVGV MARKYHLDIYANEKTWQLMDECNMLGKLDASQKHIFQDKVLTFGDVD 120

Query: 121 IESFGVSHDAVD PQFYRMKDDKSFVMLTDTGYVSDRMAGLIENADGYLIESNHDIEILR 180
 IESFGVSHDA+DPQFYR+MKD+KSFVMLTDTGYVSDRM G+IENADGYLIESNHDIEILR
 Sbjct: 121 IESFGVSHDAIDPQFYRIMKDNKSFVMLTDTGYVSDRMTGIIENADGYLIESNHDIEILR 180

Query: 181 SGSYPWTLKQRI LSKGHLSDGSETMIRTIGNRTKHIYLGHL SKENNIKELAHMTMEN 240
 SGSYPW+LKQRI LSKGHLSDG+ MIR++G TK IYLGHL SKENNIKELAHMTM N
 Sbjct: 181 SGSYPW LSKQRI LSKGHLSDGAGAMIRSLGYNTKKIYLGHL SKENNIKELAHMTMVN 240

Query: 241 NLMRADFGVGTDFSVHDTSPDSATPLTRI 269
 L AD VGTDF+VHDTSPD+A PLT I
 Sbjct: 241 QLAMADLAVGTDFTVHDTSPDTACPLTDI 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 362

A DNA sequence (GBSx0393) was identified in *S.agalactiae* <SEQ ID 1175> which encodes the amino acid sequence <SEQ ID 1176>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-11.94 Transmembrane 15 - 31 (5 - 34)

----- Final Results -----
 bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1177> which encodes the amino acid sequence <SEQ ID 1178>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-458-

An alignment of the GAS and GBS proteins is shown below:

Identities = 335/443 (75%), Positives = 392/443 (87%)

```

5   Query: 7  NIRSFELALLFLLVFVAVYFVYLAVRDFKMSKNIRLLNWKVRDLIAGNYSDSILIQQDAD 66
      N+ +FELA+L LLVFVA YF++LAVRD++ ++ IR+++ K+RDLI G Y+D I + D +
      Sbjct: 8  NLSTFELAILILLVFVAFYFIHLAVRDYRNARIIRMMSHKIRDLINGRYTDTIDEKADIE 67

10  Query: 67  LVELGESLNDLSDVFRMAHDNLEQEKRLASILTYMTDGVLATDRSGKIVMINETAQQQF 126
      L+EL + LNDLSDVFR+ H+NL QEKNRLASIL YM+DGVLATDRSGKI+MINETA++Q
      Sbjct: 68  LMELSDQLNDLSDVFRLLTHENLAQEKNRLASILAYMSDGVLATDRSGKIIMINETARKQL 127

      Query: 127 NLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRRDENGFEVTLRIRFALNRRESG 186
      NL+ +EAL NI D+L + Y+++DLVSKTP V +N R++ GEFV+LR+RFALNRRESG
15  Sbjct: 128 NLSKEEALKKNITDLEGDTSYTYRDLVSKTPVVTVNSRNDMGFVSLRLRFALNRRESG 187

      Query: 187 FISGLVAVSHDATEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEEVAPSF 246
      FISGLV V HD TEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGAL E++APSF
20  Sbjct: 188 FISGLVVVLHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALKEDIAPSF 247

      Query: 247 IKVSLDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSILNRFQIRNQKTVTG 306
      IKVSLDETNRMMRMISDLL+LSRIDN+VT L VEMTNFTAF+TSILNRFD ++NQ T TG
      Sbjct: 248 IKVSLDETNRMMRMISDLLNLSRIDNQVTQLAVEMTNFTAFITSILNRFDLVKNQHTGTG 307

25  Query: 307 KVEIVRDYPLKSIWVEIDTDKMTQVIDNILNNAVKYSPDGGKITVNLRTTKTQMILSIS 366
      KVEIVRDYP+ S+W+EID DKMTQVI+NILNNA+KYSPDGGKITV ++TT TQ+I+SIS
      Sbjct: 308 KVEIVRDYPITSVWIEIDNDKMTQVIENILNNAIKYSPDGGKITVRMKTDTQLIISIS 367

30  Query: 367 DQGLGIPKDLPLIFDRFYRVDKARSRQGGTGLGLSIAKEIVKQHGFIWAKSEYKGKS 426
      DQGLGIPK DLPLIFDRFYRVDKARSR QGGTGLGL+IAKEI+KQH GFIWAKS+YKGKS
      Sbjct: 368 DQGLGIPKTDLPLIFDRFYRVDKARSRAQGGTGLGLAIAKEIHKQHGFIVAKSDYKGKS 427

      Query: 427 TFTIVLPYDKDAVTYEEWEDVED 449
      TFTIVLPY+KDA YEEWE+ D
35  Sbjct: 428 TFTIVLPYEKDAAYEEWEEDVD 450

```

A related GBS gene <SEQ ID 8561> and protein <SEQ ID 8562> were also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1  Crend: 8
      McG: Discrim Score: 8.59
      GvH: Signal Score (-7.5): -3.38
          Possible site: 26
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 1 value: -11.94 threshold: 0.0
45  INTEGRAL Likelihood = -11.94 Transmembrane 15 - 31 ( 5 - 34)
      PERIPHERAL Likelihood = 8.27 178
      modified ALOM score: 2.89

      *** Reasoning Step: 3

50  ----- Final Results -----
          bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
55

```

The protein has homology with the following sequences in the databases:

```

67.5/83.5% over 439aa
          Streptococcus pneumoniae
60  GP|5830524| histidine kinase Insert characterized

      ORF01458(331 - 1647 of 1947)
      GP|5830524|emb|CAB54569.1|AJ006392(10 - 449 of 449) histidine kinase {Streptococcus
      pneumoniae}
      %Match = 45.6

```

```
%Identity = 67.5  %Similarity = 83.4
Matches = 297  Mismatches = 70  Conservative Sub.s = 70
```

45

50 Purified Thio-GBS41-His is shown in Figure 244, lane 10.

Example 363

Possible site: 60
>>> Seems to have no N-terminal signal sequence

-460-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2754(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1181> which encodes the amino acid sequence <SEQ ID 1182>. Analysis of this protein sequence reveals the following:

Possible site: 60

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2754(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 205/236 (86%), Positives = 221/236 (92%)

20 Query: 1 MKKILIVDDEKPISDIIFNLTKEGYETATAFDGREALVQYAEFQPDLIILDLMLPELDG 60
 MKKILIVDDEKPISDIIFNLTKEGY+ TAFDGREA+ + E +PDLIILDLMLPELDG
 Sbjct: 1 MKKILIVDDEKPISDIIFNLTKEGYDIVTAFDGREAVTIFEEKEKPDLIILDLMLPELDG 60

25 Query: 61 LEVAKEVRKTSHIPIIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT 120
 LEVAKE+RKTSH+PIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT
 Sbjct: 61 LEVAKEIRKTSHVPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRT 120

30 Query: 121 NIETAVAEESAQNASSDITIGELQILPDAFIKKRGEIEIETHREFELHHLATHIGQVM 180
 IETAVAE+ + + +TIG LQILPDAF+AKK G+E+ELTHREFELHHLA H+GQVM
 Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVLEIETHREFELHHLANHMGMQVM 180

35 Query: 181 TREHLLLETWGYDYFGDVRTVDVTVRRLREKIEDTPGRPEYILTRRGVGYMKSYE 236
 TREHLL E VWGYDYFGDVRTVDVTVRRLREKIEDTP RPEYILTRRGVGYMKSY+
 Sbjct: 181 TREHLL EIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGYMKSYD 236

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 364

40 A DNA sequence (GBSx0395) was identified in *S.galactiae* <SEQ ID 1183> which encodes the amino acid sequence <SEQ ID 1184>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 43

 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3791(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14701 GB:Z99118 glutamine ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 149/244 (61%), Positives = 200/244 (81%), Gaps = 2/244 (0%)

55 Query: 3 LISYKNVKNKYGDYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDDGSLVV 62
 +I+++NVNK+YGD+H L+QINL+IE G+VVV++GPSGSGKSTL+R +N LESI++G L V
 Sbjct: 1 MITFQNVNKHGDFHVLKQINLQIEKGEVVVIGGPSGSGKSTLLRCINRLESINEGVLT V 60

-461-

Query: 63 NGHELANISSKELVNLKKEVGMVFQHFNLPHKTVLENITLAPIKVLKQSKKEAMEIAEK 122
 NG + N ++ +R+ +GMVFQHF+LYPHKTVL+NI LAP+KVL+QS ++A E A
 Sbjct: 61 NGTAI-NDRKTDINQVRQNIQGMVFQHFHLYPHKTVLQNIIMLAPVKVLRQSPQAKETARY 119

5 Query: 123 YLKFEVNMWERKDSYPSMLSGGQKQRIATARGLAMHPKLLLFDEPTSALDPETIGDVLSVM 182
 YL+ V + ++ D+YPS LSGGQ+QR+AIARGLAM P+++LFDEPTSALDPE IG+VL VM
 Sbjct: 120 YLEKVGIPDKADAYPSQLSGGQQORVAIARGLAMKPEVMLFDEPTSALDPEMIGEVLDM 179

10 Query: 183 OKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFDNPREPRAKQFLSN 242
 + LA +GM MVVVTHEMGFA+EVADRI+F+ +G+IL + +F+ NP+E RA+ FLS
 Sbjct: 180 KTLAKEGTMVVVTHEMGFAKEVADRIVFIDEKGILLEAVPA-EFYANPKEERARLFLSR 238

Query: 243 IINH 246
 I+NH
 15 Sbjct: 239 ILNH 242

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1185> which encodes the amino acid sequence <SEQ ID 1186>. Analysis of this protein sequence reveals the following:

Possible site: 51
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3763 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 131/243 (53%), Positives = 179/243 (72%), Gaps = 2/243 (0%)

30 Query: 2 SLISYKNVKNKYGYDYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDGSLV 61
 ++IS K+++KYYG L+ I+L+I PG+VVV++GPSGSGKSTL+RTMN LE G +
 Sbjct: 5 AIISIKDLHKYGHNEVLKGIDLDIMPGEVVVIIGPSGSGKSTLLRTMNLLEVPTKGQIR 64

35 Query: 62 VNGHELANISSKELVNLKKEVGMVFQHFNLPHKTVLENITLAPIKVLKQSKKEAMEIAE 121
 G ++ + ++ ++R+++GMVFQ FNL+P+ T+LENITL+PIK +K EA + A
 Sbjct: 65 PEGIDITD-KKNDIFSMREKMGVMVFQFNLFPNMILENITLSPIKTKGMAKAEADKTAL 123

40 Query: 122 KYLKFEVNMWERKDSYPSMLSGGQKQRIATARGLAMHPKLLLFDEPTSALDPETIGDVLSV 181
 L V + E+ +YP+ LSGGQ+QRIATARGLAM P +LLFDEPTSALDPE +G+VL+V
 Sbjct: 124 SLLDKVGLSEKAKAYPASLSGGQQRIATARGLAMDPDVLVLFDEPTSALDPEMVGVLAV 183

45 Query: 182 MQKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFDNPREPRAKQFLS 241
 MQ LA GM MV+VTHEMGFA+EVADR++FM DG ++V+ FD +E R K FLS
 Sbjct: 184 MQDLAKSGMTMVIIVTHEMGFAKEVADRVFM-DGGVIVEEGSPNQLFDLTKEERTKDFLS 242

Query: 242 NII 244
 ++
 Sbjct: 243 RVL 245

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 365

A DNA sequence (GBSx0396) was identified in *S.agalactiae* <SEQ ID 1187> which encodes the amino acid sequence <SEQ ID 1188>. This protein is predicted to be glutamine-binding. Analysis of this protein
 55 sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-462-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73178 GB:AL139076 probable ABC-type amino-acid transporter
 periplasmic solute-binding protein [Campylobacter
 jejuni]
 Identities = 99/240 (41%), Positives = 141/240 (58%), Gaps = 3/240 (1%)

10 Query: 1 MLRRKRLTFYLLSCIFIFLLFYPNSTSANQLSEIKKSGVLKVGKQDVPNFGYYNAETNQ 60
 M+ RK L + + + F + + +L IK G L VGVK DVP++ + T +
 Sbjct: 1 MVFRKSLKLAVFALGACVAFSNANAAEGKLESIKSGQLIVGVKNDVPHYALLDQATGE 60

15 Query: 61 YEGMEIDIAKKIAXSL---GVKPVFVPTTAQTREPLMDNGQIDILIATYTTITPERKANYN 117
 +G E+D+AK +AKS+ K V A+TR PL+DNG +D +IAT+TITPERK YN
 Sbjct: 61 IKGFVDVAKLLAKSILGDDKKIKLVAVNAKTRGPLLDNGSVDAVIATFTTITPERKRIYN 120

20 Query: 118 ISKAYYHDEIGFLVRKNSHIKTIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLG 177
 S+ YY D IG LV K K++ ++ G +IGVAQ ATTK + + AK+ + +++
 Sbjct: 121 FSEPPYQDAIGLLVLKEKQYKSLADMKGANIGVAQAATTKAIGEAACKIGIDVKFSEFP 180

25 Query: 178 SFPELAISLYANRIDAFSVDKSILSGYLSPHTTILKEGFNTQEYGIATSKQDKVLIPYVN 237
 +P + +L A R+DAFSVDKSIL GY+ + IL + F Q YGI T K D YV+
 Sbjct: 181 DYPSIKAALDAKRVDAFSVDKSILLGYVDDKSEILPDSFEPQSYGIVTKKDDPAFAKYVD 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1189> which encodes the amino acid sequence <SEQ ID 1190>. Analysis of this protein sequence reveals the following:

Possible site: 30
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.16 Transmembrane 17 - 33 (15 - 35)

----- Final Results -----
 35 bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9097> which encodes the amino acid sequence <SEQ ID 9098>. Analysis of this protein sequence reveals the following:

40 >>> May be a lipoprotein

----- Final Results -----
 45 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 66/251 (26%), Positives = 111/251 (43%), Gaps = 27/251 (10%)

50 Query: 23 PNSTSANQLSEIKKSGVLKVGKQDVPNFGYYNAETNQYEGMEIDIAKKIAXSLGVKPVF 82
 P+ + + IK+ GVLKV +YN + N+ G E+D+ K+I K L +K F
 Sbjct: 34 PHQSQSSWDTIKEKGVKLVATPGTYQPTSFYN-DNNELVGYEVDMMVKEIGKRLNIKVKF 92

55 Query: 83 VPTTAQTREPLMDNGQIDILIATYTTITPERKANYNISKAYYHDEIGFLVR---KNSHIK 138
 V T +D+G++DI + + ITP+R+ YNIS Y + G +VR N K
 Sbjct: 93 VETGFDQAFTSVDSGRVDISLNNFDITPKRQKKYNISTPYKYGVGGMIVRADGSSNIAKK 152

60 Query: 139 TIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLGSPPELAISLYANRI----- 191
 + + GK A G +K A+L ++ + +Y N +
 Sbjct: 153 DLSDWKGKKAAGASGTEYMKVAQKQG-----ABLVTYDNTGVDVYLNDVANGRTDF 203

Query: 192 --DAFSVDKSILSGYLSPHTTILKE---GFNTQEYGIATSKQDKVLIPYVNKLVSWEK 245

-463-

++ K + LS + + + +N E GI +K+D L ++ ++ K
 Sbjct: 204 IPNDYPAQKLFDVYMLSQNPNLNVKMSDVQYNPTEQGIVMNKKDDSLKKKIDAVIKDMIK 263

Query: 246 DGSLKHIYQKF 256

DGSLK I + +

Sbjct: 264 DGSLKKISETY 274

SEQ ID 1188 (GBS136) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 5; MW 29.9kDa).

- 10 The GBS136-His fusion product was purified (Figure 200, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 284), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 366

A DNA sequence (GBSx0397) was identified in *S.agalactiae* <SEQ ID 1191> which encodes the amino acid sequence <SEQ ID 1192>. This protein is predicted to be integral membrane. Analysis of this protein sequence reveals the following:

20 Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.34 Transmembrane 32 - 48 (27 - 55)
 INTEGRAL Likelihood = -5.04 Transmembrane 200 - 216 (196 - 219)
 INTEGRAL Likelihood = -3.13 Transmembrane 93 - 109 (93 - 113)
 INTEGRAL Likelihood = -2.02 Transmembrane 74 - 90 (74 - 92)
 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73177 GB:AL139076 putative ABC-type amino-acid transporter
 permease protein [Campylobacter jejuni]
 Identities = 112/226 (49%), Positives = 160/226 (70%), Gaps = 3/226 (1%)
 35 Query: 5 NISPF AISRWGAFFNHFDFKGFLYTLGISFGALLALILGILSGGLSTSKSKVGKLIS 64
 +ISPPA+ ++ ++ D F GF+YTL +S ALL+A I G + G ++TS+ K+ + +
 Sbjct: 25 SISPF AVWKFLDALDNKDAFINGFIYTLLEVSIALLIATIFGTIGGVMATSRFKIIRAYT 84
 40 Query: 65 RIYVEVFQNTPLLVQMVFVYGLAISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGIE 124
 RIYVE+QFN PL++Q+ F++Y L ++ + + F VL VG YHGAY+SEV+RSGI
 Sbjct: 85 RIYVELFQNVPLVIQIFFLFYALPVLG---IRLDIFTIGVLGVGAYHGAYVSEVVRSGIL 141
 45 Query: 125 AVPKGQTEAALAQGF TANQTMQLIILPQAVRTILPPMTNQVNLIKNTSTVAIISGADIM 184
 AVP+GQ EA+ +QGFT Q M+ II+PQ +R ILPPMTNQ+VNLIKNTS + I+ GA++M
 Sbjct: 142 AVPRQGFEASASQGFTYIQQMRYIIVPQTIRIILPPMTNQMVNLIKNTSVLLIVGGAELM 201
 Query: 185 FVAKAWAYDTTNYIPAFAGAAIFYFVICFPLASWARKQEELNKKTY 230
 A ++A D NY PA+ AA+ YF+IC+PLA +A+ E KK +
 50 Sbjct: 202 HSADSYADYGNYPAYIFAAVLYFIICPLAYFAKAYENKLKKAH 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1193> which encodes the amino acid sequence <SEQ ID 1194>. Analysis of this protein sequence reveals the following:

55 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

-464-

INTEGRAL Likelihood = -6.26 Transmembrane 307 - 323 (303 - 327)
 INTEGRAL Likelihood = -5.89 Transmembrane 485 - 501 (479 - 502)
 INTEGRAL Likelihood = -1.12 Transmembrane 375 - 391 (375 - 391)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
 [Synechocystis sp.]
 Identities = 146/532 (27%), Positives = 244/532 (45%), Gaps = 59/532 (11%)

15 Query: 6 YMKKLILSCLVALALLFGGMSRAQANQYLRVGMEAAAYAPFNWTQDDASNGAVPIEGTSQY 65
 Y L L L+A+A+ + Q + V E + PF T E T Q
 Sbjct: 16 YYLLALGVLLAIAIPLLPFSQVSRQTIIVATEPTFPFPEMTD-----EATGQL 65

20 Query: 66 ANGVDVQVAKKAKAMNKELLVVKTSWTGLIPALTSGKIDMIAAGMSPTKERRNEISFSN 125
 G+DV + + + +A + + + G+IPAL S + + + T ER +SFS+
 Sbjct: 66 T-GFDVDLIQAIGEAQVTVDIQGYPPDGIIIPALQSNVTGAAISAITITPERAQSVSFSS 124

25 Query: 126 SSYTSQPVLVVTANGKYADATSLKDFSGAKVTAQQGVWHVNLITQLKGAKLQTPMGDFSQ 185
 + S VL + +LKD G ++ G + T + GAK+ T +
 Sbjct: 125 PYFKS--VLAIAVDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPCAKV-TNFDITS 181

30 Query: 186 MRQALTSVIDAYISERPEAMTAEADSRLKMITLKKGFAVAESDAAIAGVMKKNDDRMA 245
 Q L +G DA I++RP + A D+ L+ + + +E IA+ + +
 Sbjct: 182 ALQELVNGNADAVINDRPVLLYA-IKDAGLRNVKISADVG-SEDYVGIAMPLAPPE--- 236

35 Query: 246 TVNQVLEGFSQTDRLMDDMVTKQPVKEKAEDAKASFLGQMWAIFKGN----- 294
 +NQ E +Q +++++ EK + FL + G
 Sbjct: 237 -INQFREVLNQ-GLFQIIENGTYNAIYBKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERS 294

40 Query: 295 -----WKQFLRGTMILLISMVGTITGLFIGLLIGIFRTAPKAKHKVAALGQK 342
 ++ +G+ +T+L++ GL G + I + K
 Sbjct: 295 QANPNDNFLITLFRNLFKGSILTVLLTAFSVFFGLIGGTGVAIALISDI-----K 344

45 Query: 343 LFGWLLTIYIEIFRGTPMIVQSMVIYYGTAQAF-----GISIDRTLAAIFIVSINTGAYM 397
 + IY+E FRGTPM+VQ +IY+G F GI+IDR AAI +S+N AY+
 Sbjct: 345 PLQLIFRIYVEFFRGTPMLVQLFIYFGLPALFKEIGLGTIDRFPAIIALSLNVAAYL 404

50 Query: 398 SEIVRGGIFAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNIPATGNEFVINIKDTSVL 457
 +EI+RGGI ++D+GQ++A +LG + QTM++++ PQ R ILP GNEF+ IKDTS+
 Sbjct: 405 AEIIRGGIQSIDQGWEACESLGSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSIT 464

Query: 458 NVISVVELYFSGNTVATQTYQYFQFTTIIAIYFVLTFTVTRILRYIERFD 509
 VI EL+ G + TY+ F+ + +A++Y +LT + + +++E D
 Sbjct: 465 AVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTITISSFVFKWLENYMD 516

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/210 (39%), Positives = 113/210 (53%), Gaps = 12/210 (5%)

55 Query: 14 WGAFFNHFDLFFPKGFLYTLGISFGALLLALILGILSGGLSTS---KSKVGKL-----I 63
 W F ++ F+G TL IS + L +G+L G T+ K KV L +
 Sbjct: 288 WAIFKGNWKQFLRGTMILLISMVGTITGLFIGLLIGIFRTAPKAKHKVAALGQKLFQWL 347

60 Query: 64 SRIYVEVFQNTPLLVQMFVYVGLAISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGI 123
 IY+E+F+ TP++VQ + +YYG A + I A+ V + GAY+SE++R GI
 Sbjct: 348 LTIYIEIFRGTPMIVQSMVIYYGTAQAF--ISIDRTLAAIFIVSINTGAYMSEIVRGGI 405

65 Query: 124 EAVPKGQTEAALAQGFTANQTMQLIILPQAVRTILPPMINQVNVNLIKNTSTVAIISGADI 183
 AV KGQ +AA A GFT QIM+ I+LPQ VR ILP N+ V IK+TS + +IS ++
 Sbjct: 406 FAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNIPATGNEFVINIKDTSVLNVISVVEL 465

Query: 184 MFVAKAWAYDTTINYIPAFAGAAIFYFVICF 213

-465-

F A T Y F AI YFV+ F
 Sbjct: 466 YFSGNTVATQTYQYFQTFTTIIAIYFVLTF 495

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 367

A DNA sequence (GBSx0398) was identified in *S.agalactiae* <SEQ ID 1195> which encodes the amino acid sequence <SEQ ID 1196>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.95 Transmembrane 25 - 41 (16 - 42)
 INTEGRAL Likelihood = -3.61 Transmembrane 66 - 82 (65 - 86)
 INTEGRAL Likelihood = -2.44 Transmembrane 184 - 200 (182 - 201)
 INTEGRAL Likelihood = -0.59 Transmembrane 119 - 135 (119 - 135)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14704 GB:Z99118 glutamine ABC transporter (integral membrane protein) [Bacillus subtilis]
 Identities = 84/206 (40%), Positives = 129/206 (61%), Gaps = 6/206 (2%)
 Query: 10 ILFLQGGFGLTLYISFISILLSMFFGTLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWI 69
 + FL GF +TLY++FISI+LS FFG + +R +K P+ + ++ +E +RN+P LL I
 Sbjct: 12 LAFLDWGDFLVTLTVAFISIIISFFGLIAGTLRYAKVPVLSQLIAVLVETIRNLPPLLLII 71
 Query: 70 FIIFLVF-----QMKSVSAGITSFTIFTSAALAEIIRGGLNGVDKQGQTEAGLSQGFTYLO 124
 F F +++ +A IT+ TIF SA L+EIIR GL +DKGQ EA S G +Y Q
 Sbjct: 72 FTTFFALPEIGIKLEITAAAITALTIFESAMLEIIRSGLSIDKGQIEAARSSGLSYTQ 131
 Query: 125 VFIIIFPQAFRKMPLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAGQVFT 184
 I+ PQA R+M+P I+SQF++++KDTSL VIA+ E+ +QI+ G+ + F
 Sbjct: 132 TLFFIVMPQALRRMVPIVSQFISLLKDTSLAV-VIALPELIHNAQIINGQSADGSYFFP 190
 Query: 185 LYAIITAVYFITNFIISFSRKLKSKR 210
 ++ + +YF N+ +S +R+L R
 Sbjct: 191 IFLAALMYFAVNYSLSLAARRLEVR 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1197> which encodes the amino acid sequence <SEQ ID 1198>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.51 Transmembrane 529 - 545 (517 - 551)
 INTEGRAL Likelihood = -10.30 Transmembrane 697 - 713 (693 - 719)
 INTEGRAL Likelihood = -4.41 Transmembrane 560 - 576 (555 - 585)
 INTEGRAL Likelihood = -0.32 Transmembrane 662 - 678 (662 - 678)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein

-466-

[Synechocystis sp.]

Identities = 153/475 (32%), Positives = 251/475 (52%), Gaps = 27/475 (5%)

5 Query: 273 IVSDSSFAFFEFQN-GKGKYGIDIELIKAIKQGGFKIETANPGFDAALNAVQSSQADG 331
 + ++ +F PFE + G+ G D++LI+AI + ++I FD + A+QS+
 Sbjct: 46 VATEPTFPFEMTDEATGQLTGFDVDLIQAIGEAAQVTDIQGYPFDDGIIPALQSNVTGA 105

10 Query: 332 VIAGATITDARKAIFDFSDPYTSTNIILAVKAGKN-IKNYEDLDRKTGAKNGTSSYSWL 390
 I+ TIT R FS PY+ S + +AV+ G + IKN +DL+ K + GT+ + +
 Sbjct: 106 AISAITITPERAQSVSFSSPYFKSVLAIAVQDNDTIKNLKDLEGKRLAVAIGTTG-AMV 164

15 Query: 391 KENAPKYGYNVKAFDDGSSMYDSLNSGSDAIDMDEAVLKYAISQG--RRFETPLEGIST 448
 N P G V FD +S L +G+ DA+++D VL YAI R + + S
 Sbjct: 165 AINVP--GAKVTNFDSTISALQELVNGNADAVINDRPVLLYAIKDAGLRNVKISADVGSE 222

20 Query: 449 GEVGFVAVKKGSTNPELI---EMFNNGLAALKKSGQYDDIIDKYLDSKKA-----ATPSEKG 500
 G A+ E+ E+ N GL + ++G Y+ I +K+ K PS G
 Sbjct: 223 DYYGIAMPLAPPGEINQTRVLNQGLFQIIENGTYNAIEYKWFGEKNPPFLPLVAPSLVG 282

25 Query: 501 -----ADESTISGLLSNNYKQLLAGLGTTLSTLISFAIAIIIGIIFGMMAVSP 549
 + + L ++ L G T+ LT S +I G + +S
 Sbjct: 283 KVGTAQSLTERSQANPDNFLTITLERNLFKGSILTLLTAFSVFGLIGGTGVAIALISD 342

30 Query: 550 TKSLRLISTVFDVVRGIPLMIVAFAFIWGVPNLIESMTGHQSPINDFLAATIALSLNGG 609
 K L+LI ++V+ RG P+++ I++G+P L + + G I+ F AA IALSLN
 Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKBI-GLGITIDRFPAALIALSLNVA 401

35 Query: 610 AYIAEIVRGGIEAVPAGQMEASRSLGLSYGTTMRKVILPQAVKMLLPNFINQFVISLKD 669
 AY+AEI+RGGI+++ GQ EA SLG+S TM++VI PQA + +LP N+F+ +KDT
 Sbjct: 402 AYLAETIRGGIQSIDQGWEEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDT 461

40 Query: 670 TIVSAIGLVELFQTGKIIIIARNYQSRMYAILAIYLMILLTRLAKRLEKRLN 724
 ++ + IG ELF+ G++I+A Y++F +Y +A++YL++ + + + K LE ++
 Sbjct: 462 SLTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTISSEFVKWLENYMD 516
 Identities = 68/247 (27%), Positives = 106/247 (42%), Gaps = 11/247 (4%)

45 Query: 7 VLLLAIMSIFLTCNIASETIAIVSDTAYAPFEFKD--SDQIYKIDVDIINEVAKRQSW 64
 VLL + + + S +TI + ++ + PFE D + Q+ G DVD+I + +
 Sbjct: 24 VLLAIAIPLLPAPFSQVSRQTIIVATEPTFPFEMTDEATGQL-TGFDVDLIQAIGEAAQV 82

50 Query: 65 DFSMSFPGFDAAVNAVQSGQASALMAGTTITNARKKVFHFSEFYDTKIVIATRKA-AT 123
 + FD + A+QS A ++ TIT R + FS PY+ + + IA + N I
 Sbjct: 83 TVDIQGYPFDDGIIPALQSNVTGAASAITITPERAQSVSFSSPYFKSVLAIAVQDNDTI 142

55 Query: 124 KKYSDLKGKTVGVKNGTAAQAFLLNNYKKKYDYTVKTFDTGDLMYNSLSAGSIAAVMDDEA 183
 K DL+GK + V GT N V FD+ L G+ AV++D
 Sbjct: 143 KNLKDLEGKRLAVAIGTTGAMVATNVP---GAKVTNFDSTISALQELVNGNADAVINDRP 199

Query: 184 VIQYAIS----QNODIAINMKGEPIGSFGFAVKKGSGYDYLVDNFNTALKAMKADCTYQA 239
 V+ YAI +N I+ ++ E + + N L + +GTY A
 Sbjct: 200 VLLYAIKADAGLRNVKISADVGSEDDYGIAMPLAPPGEINQTRVLNQGLFQIIENGTYNA 259

Query: 240 IMTKWLG 246
 I KW G
 Sbjct: 260 IYKWF 266

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 113/210 (53%), Gaps = 16/210 (7%)

60 Query: 13 LLQGFGLTYISFISILLSMFFGTLALMRNSKNPIWKLIASIIYEFVRNVNLLWIFII 72
 LL G G TL ++ IS +++ G + +M S +LI+++++ VR +P ++ I
 Sbjct: 517 LLAGLGTTLSTLISFAIAIIIGIIFGMMAVSPTKSLRLISTVFDVVRGIPLMIVAFAFI 576

65 Query: 73 F-----LVFQMKSVSAGITSFTIFT-----SAALAEIIRGGGLNGVDKQTEAGLSQGF 120
 F L+ M + I F T A +AEI+RGG+ V GQ EA S G
 Sbjct: 577 FWGVNPLIESMTGHQSPINDFLAATIALSLNGGAYIAEIVRGGIEAVPAGQMEASRSLGL 636

-467-

Query: 121 TYLQVFIIIIFPQAFRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAG 180
 +Y +I PQA + MLP I+QFV +KDT+++ S I + E+F +I++ R +
 Sbjct: 637 SYGTTMRKVILPQAVKLMLENFINQFVVISLKDITIV-SAIGLVELFQTGKIIARNY--- 692

5 Query: 181 QVFTLYAIITAVYFITNFISSFSRKLSKR 210
 Q F +YAI+ +Y I +++ ++L KR
 Sbjct: 693 QSFRMYAILAIIVLIMIILLTRLAKRLEKR 722

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 368

A DNA sequence (GBSx0399) was identified in *S.agalactiae* <SEQ ID 1199> which encodes the amino
 acid sequence <SEQ ID 1200>. Analysis of this protein sequence reveals the following:

Possible site: 39
 15 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.21 Transmembrane 7 - 23 (1 - 30)

----- Final Results -----
 20 bacterial membrane --- Certainty=0.5883(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA04094 GB:AP001508 unknown conserved protein in B. subtilis
 25 [Bacillus halodurans]
 Identities = 43/157 (27%), Positives = 83/157 (52%), Gaps = 9/157 (5%)

Query: 26 YQSOFQKTTNQALAIAYKDAKVAKK--DVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHV 83
 +Q++ N+ L +A ++ + + + +K+ +N R YEIE EY + +
 30 Sbjct: 38 HQAESVSADNEGLTLAEASDIALERAGNGVVTEAEKDRDNGRVVYEIEVKNDDEYDFKI 97

Query: 84 DVKTGQILERMDNNGFSKSTSQSSSSSQSKSHKISQEEAKKIAFKDANIEESEVSNLKI 143
 D +TG+IL+ + SK SSS ++ IS +EAK+IA K+ + ++ ++++
 35 Sbjct: 98 DQQTGEILKEKQEQKSGKPREGHSSSKGSEA-VISMDEAKEIALKEVS---GKIDDIEL 153

Query: 144 KEEIENGKSVYDIDF-VDLKKNKNEVDYQIDAETGKII 179
 E ENG VY+++ D ++V +DA TG ++
 Sbjct: 154 --ERENGLVYEVEIESDHYDDDDVTYVDAMTGNVL 188

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1201> which encodes the amino acid
 sequence <SEQ ID 1202>. Analysis of this protein sequence reveals the following:

Possible site: 57
 45 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.15 Transmembrane 42 - 58 (41 - 60)

----- Final Results -----
 50 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 37/96 (38%), Positives = 63/96 (65%), Gaps = 5/96 (5%)
 55 Query: 94 DMDNNGFSKSTSQSSSSSQSKSHKISQEEAKKIAFKDANIEESEVSNLKIKEEENGKSV 153
 DMD+ +Q +S + K K+S+++AK IA KDA++ E++ L + ++ E+GK+V
 Sbjct: 59 DMDDKD-DHMDNQPKTSQTSKKVLSKEDKAKSIALKDASVTEADAQMLSVTQDNEDGKAV 117

-469-

```

762      792      822      852      882      912      942      972
EEIENGKSVYDIDFVLDLKNKNEVDYQIDAETGKI IERSRDHMD*FK*DIKKRRSRKPSF*LLSLLPTF*KFT*KT*DD
|:  :|  :|:::  :|  |:::|  |:|:|::  |
ED--DGAYIYEME-IQTKQGTETETEFISAKDGRIIKQEIDD
5      140      150      160      170

```

SEQ ID 8564 (GBS37) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 4; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 10; MW 47kDa).

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 369

A DNA sequence (GBSx0400) was identified in *S.galactiae* <SEQ ID 1203> which encodes the amino acid sequence <SEQ ID 1204>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 53
>>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.1499(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9709> which encodes amino acid sequence <SEQ ID 9710> was also identified.

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1205> which encodes the amino acid sequence <SEQ ID 1206>. Analysis of this protein sequence reveals the following:

- Possible site: 42
>>> Seems to have no N-terminal signal sequence
- 30 ----- Final Results -----
- bacterial cytoplasm --- Certainty=0.2808(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 35

An alignment of the GAS and GBS proteins is shown below:

- Identities = 128/297 (43%), Positives = 180/297 (60%), Gaps = 9/297 (3%)
- 40 Query: 54 IDDIKVGSPIFKYFWT-SLSLQAPLKALEFVLEQAKMPTELSCGELSETQYLVAQFSDELA 112
I D ++GSP F W Q+ + L F+L+ +MP ELSG+L ETQ L+ +F L
Sbjct: 46 IIDNRLGSPTFWVIWPIEKENQSAKQLLTFLDLVEMPFELSGQLHETQTLTTRFHPSL 105
- Query: 113 PHDDFWIALSQVIYDSFPGNSLAEDTVLNRKLHQFRYLSSQQAQYVRRYFKDVGMTDRD 172
P FW L+ ++ +FPG +L++ L ++LHQFRY+ISSQQAQ +R ++K + MTD
45 Sbjct: 106 PDHMFWKELASLVDQAFPGKTLQAGELEKRLHQFRYVISSQQAQSIRNHYKMIEMTDAQ 165
- Query: 173 ALVNYL-----SCL-REPDSIAYYESARLHNKRRRNGEIFGFPDDEPVINSKLLISFHTE 226
AL +L CL R+ +SARLHNK R FP E N K+L+ FHTE
50 Sbjct: 166 ALALFLRSKKGPCLRQAPDYTLMD SARLHNKLR FEDNKFVFP SQEVSYNIKVLLWFHTE 225
- Query: 227 FIIDDKGNFLNEIDA EVITRNGIINGASFNYAFKNTRHKELDVPVK-LDPKFRNDMTR 285
F +D G FLNE+DAEV+T GI+NGASFNY + RH +LDVDP+ DP+FR D +
Sbjct: 226 FTLDSTGFFLNEVD AEVTEKGIVNGASFNYG-TDGP RHWDLDVDPISHHDPQFRDITLK 284

-470-

Query: 286 GYRSPNLSRRKWNFFKEDYDCSYFNKKGYAFGRRSKQSVKQVLYKKAVQKMR 342
 G+RSP R+WF +++D+ SYFN KG +A+ +S+ V K K K+ + ++
 Sbjct: 285 GFRSPKRVFRQWFRQKDDFMFSYFNAKGLFAYHNKSSFARVKKSAKQFKRQIHPIK 341

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 370

- A DNA sequence (GBSx0401) was identified in *S.agalactiae* <SEQ ID 1207> which encodes the amino acid sequence <SEQ ID 1208>. This protein is predicted to be similar to two-component response regulator [YcbM] (ompr-likeprotei). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3129(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 20 >GP:CAA55264 GB:X78502 gtcR [Brevibacillus brevis]
 Identities = 99/228 (43%), Positives = 149/228 (64%), Gaps = 3/228 (1%)
- Query: 2 RTVLVVQGDDETIELLSYLEGALYKVMASDGEEAFSLFQQHQIDLAIIDITLPKIDGY 61
 +T+L+ + E IELL+ +LE Y+++ A DGE+A++ +QH +DLAIIIDI +P +DG+
 25 Sbjct: 3 KTLIIADDEPEIIELLKFLERESYRIIEAYDGEQAWNYIRQHPVLDLAIIDIMPALDGF 62
- Query: 62 ELTRLIRQDSQIPPIIMLAAKTTMDRILGLNIGADDFITKPFNSLEVLARINSQLRYYE 121
 +L + + + ++P+I+L+AK D D+ILGL +GADDFI+KPFN LE +ARI +QLRR +E
 30 Sbjct: 63 QLIIKRLTNEYKLPVILSAKNRSDKILGLGLGADDFISKPFNPLEAVARIQAQLRRAFE 122
- Query: 122 FNSLAKP--KNQFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMS-SPGRIYTKT 178
 FN + Q +G L L + + + +T E+++L+ M S I+TK
 Sbjct: 123 FNEPBEKAISTQSTTVGRLLTLLHTACVVYRGDETYSVTPLEYRLNLTFMQCSRTSIFTKQ 182
- 35 Query: 179 QLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGYK 226
 QL+E+ D+ TIMV IS +RDKIED + P YIKT+RG+GYK
 Sbjct: 183 QLFEQAWSETYWEDDNTIMVQISRLDKIEDQPRQPVIYIKTVRGLGYK 230

There is also homology to SEQ ID 1182:

- 40 Identities = 87/230 (37%), Positives = 144/230 (61%), Gaps = 5/230 (2%)
- Query: 1 MRTVLVVQGDDETIELLSYLEGALYKVMASDGEEAFSLFQQHQIDLAIIDITLPKIDG 60
 M+ +L+V + +++++ L Y +V A DG EA ++F++ + DL I+D+ LP++DG
 45 Sbjct: 1 MKKILIVDDEKPISDIIFNLTKEGYDIVTAFDGREAVTIFEEKPDLIILDMLPELDG 60
- Query: 61 YELTRLIRQDSQIPPIIMLAAKTTMDRILGLNIGADDFITKPFNSLEVLARINSQLRYY 120
 E+ + IR+ S +PIIML+AK ++ D+++GL IGADD+TKPF++ E+LAR+ + LRR
 Sbjct: 61 LEVAKEIRKTSHPPIIMLSAKDSEFDKVGLEIGADDYVTKPFNSRELLARVKAHLRRT 120
- 50 Query: 121 EFNSLAKPKN----QFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMSSPGRIY 175
 + +N Q + IG L++ + K+G+ ++LT EF++LH L + G++
 Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVETLHREFELLHLLANHMGQVM 180
- 55 Query: 176 TKTQLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGY 225
 T+ L E + G GD T+ V + +R+KIED P+YI T RGVGY
 Sbjct: 181 TREHLLIIVWGYDYFGDVRTVDVTVRRLEKIEDTPSRPEYILTFRGVGY 230

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-471-

Example 371

A DNA sequence (GBSx0402) was identified in *S.agalactiae* <SEQ ID 1209> which encodes the amino acid sequence <SEQ ID 1210>. This protein is predicted to be threonyl-tRNA synthetase 1 (thrS). Analysis of this protein sequence reveals the following:

```

5   Possible site: 32
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.2353(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:BAB06860 GB:AP001517 threonyl-tRNA synthetase 1 [Bacillus halodurans]
   Identities = 413/638 (64%), Positives = 506/638 (78%), Gaps = 7/638 (1%)

   Query: 1   MIKITFPDGAIREFESGITTTFEIAQSISNSLAKKALAGKFNQGLIDTTRAIEDGSIEIV 60
   MI ITFPDGA++EF G TT EIA SIS L KKALAG +G L+D IE+DG+I IV
   Sbjct: 4   MINITFPDGAIVKEFPKGTITTAETIAGSISPGKKKALAGMLDGTLLDLNTPIEQDGTITIV 63

20   Query: 61   TPDHEDALGVLRHSAHLFAQAARLFPD--LCLGVGPAIQDGFYYDTDNKSGQISNDDL 118
   TP+ ++AL VLRHS AH+ AQA KRLF D + LGVGP I+ GFYYD D ++ +DL
   Sbjct: 64   TPESDEALEVLRHSTAHVMAQALKRLFKDRNVKLGVPVIEGGFYDVMDES-LTPEDL 122

25   Query: 119  PRIIEEMKKIVKENHPCIREEISKEEALFKD--DPYKVELISEHAEDG-LTVYRQGEF 175
   P+IE+EMKKI+ EN P R +S+EEAL +++ DPYK+ELI++ ED +T+Y QGEF
   Sbjct: 123  PKIEKEMKKIIGENLPIERVVVSREEALARYEEVGDPYKIELINDLPEDETITTIYQGEF 182

30   Query: 176  VDLCRGPHVPSTGRIQVFHLLNVAGAYWRGNSDNAMQRYGTAFDCKDLKAYLKRRRE 235
   DLCRG HVPSTG+++ F LLN+AGAYWRG+S N M+QR+YGTA+F K DL +L+ EE
   Sbjct: 183  FDLCRGVHVPSTGKLKEFKLLNLAGAYWRGDSSNKMLQRIYGTAFKKADLDEHLRLLEE 242

   Query: 236  AKERDHRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELYIVDKIASGYQHVVYTP 295
   AKERDHRKLGKEL +F ++ +VGQGLP WLP GATIRR +ERYIVDK E GYQHVVYTP
35   Sbjct: 243  AKERDHRKLGKELGIFALSQKVQGGLPLWLPGATIRRIERYIVDKKEKLGQYHVYTPV 302

   Query: 296  MASVEFYKTSGHWDHYREDMFPTMDMGDEEFVLRPMNCPHHIEVYKHHVHSYRELPIRI 355
   +AS E YKTSGHWDHY++DMFPTM+M + EE VLRPMNCPHH+ VYK + SYR+LP+RI
40   Sbjct: 303  LASSELYKTSGHWDHYKDDMFPTMEM-ENEELVLRPMNCPHHMVYKTEMRSYRQLPLRI 361

   Query: 356  AELGMMHRYEKGALITGLQVRVREMTLNDAHIFVTPEQIKDEFKALNLIAEYEDFNLTD 415
   AELG+MHRYE SGA++GLQVR MTLNDAHIF P+QIKDEF++ + LI +YEDF L +
50   Sbjct: 362  AELGLMHRYEMSGAVSGLQVRGMTLNDAHIFCRPDQIKDEFVVRVRLIQAVYEDFGLKN 421

   Query: 416  YRFRLSYRDPEDKHKYYDNDEMWENAQAMLKEAMDDFGLDYFEAEGBAAFYGPKLDIQVK 475
   Y FRLSYRDPEDK KY+D+D MW AQ MLKEAMD+ L+YFEAEGBAAFYGPKLD+QV+
45   Sbjct: 422  YSFRLSYRDPEDKEKYFDDNMWNKAQGMLKEAMDELELYFEAEGBAAFYGPKLDVQVR 481

   Query: 476  TALGNEETLSTIQLDLFLPERFDLKYIGADGEEHRPIMIHRGGISTMERFTAILIETYKG 535
   TALG +ETLST+QLDFLLPERFDL Y+G DG+ HRP+++HRG +STMERF A L+E YKG
50   Sbjct: 482  TALGKDETSTVQLDFLLPERFDLTYVGEDGQPHRPVVVHRGVVSTMERFVAFLLLEEKYK 541

   Query: 536  AFPTWLAPQQVSVIPISNEAHIDYAWAVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKI 595
   AFPTWLAP QV VIP+S EAH++YA V L+ GIR E+D+R+EK+ YKIR AQ QKI
55   Sbjct: 542  AFPTWLAPVQVQVIPVSPEAHLEYAKNVQETLQQAGIRVEIDERDEKIGYKIREAQMQKI 601

   Query: 596  PYQLIVGDKEMEKAHVNRVRYGSKATETKSIEEFVESI 633
   PY L++GDKE+E VNVR+YG K + + ++EFV +
60   Sbjct: 602  PYMLVLGDKEVEANGVNVRKYGEKSSSMGLDEFVRHV 639

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1211> which encodes the amino acid sequence <SEQ ID 1212>. Analysis of this protein sequence reveals the following:

-472-

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2566(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 564/644 (87%), Positives = 608/644 (93%)

Query: 1 MIKITFPDGAIREFESGITTFEIAQSIISNSLAKKALAGKFNGLIDTTTRAIEEDGSIEIV 60
 Sbjct: 1 MIKITFPDGA+REFESG+TTF+IA+SIS SLAKKALAGKFN QLIDTTTRAIEEDGSIEIV 60

15 Query: 61 TPDHEDALGVLRHSAAHFAQAARLFPDLCLGVGPATQDGFYYDTDNKSGQISNDDLPR 120
 TPDH+DA VLRHSAAHFAQAARLFP+L LGVGPAT +GFYYDTDN QQISN+DLPR
 Sbjct: 61 TPDHKDAYEVLHSAAHFAQAARLFPNLHLGVGPATIAEGFYDDTNAEQISNEDLPR 120

20 Query: 121 IEEEMKKIVKENHPCIREBEISKEEALFKDDPYKVELISEHAEDGLTVYRQGEFVDLCR 180
 IE EM+KIV EN+PCIREE++KEEALFKDDPYKVELI+EHA GLTVYRQGEFVDLCR
 Sbjct: 121 IEAEMQKIVTBNYPCIREBEVTKEEALFKDDPYKVELINEHAGAGLTVYRQGEFVDLCR 180

25 Query: 181 GPHVPSTGRIGVFHLLNVAGAYWRGNSDNMMQRVYGTAWFDKDLKAYLKRREEAKERD 240
 GPHVPSTGRIGVFHLLNVAGAYWRGNSDN MMQR+YGTAWFDKDLKAYL R EEAKERD
 Sbjct: 181 GPHVPSTGRIGVFHLLNVAGAYWRGNSDNMMQRVYGTAWFDKDLKAYLTRLLEEAKERD 240

30 Query: 241 HRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELERYIVDKELASGYQHVVYTPFMAVE 300
 HRKLGKELDLFM++ EVGQGLPFWLP+GATIRR LERYI DKE+ASGYQHVVYTPF+ASVE
 Sbjct: 241 HRKLGKELDLFMISQEVGQGLPFWLPDGTATIRRTLERYITDKELASGYQHVVYTPFLASVE 300

35 Query: 301 FYKTSGHWDHYREDMFPTMDMGDGEFVLRPMNCPHHIEVYKHHVHSYRELPIRIAEELGM 360
 YKTSGHWDHY+EDMFP MDMGDGEFVLRPMNCPHHI+VYK+HV SYRELPIRIAEELGM
 Sbjct: 301 LYKTSGHWDHYQEDMFVMDMGDGEFVLRPMNCPHHIQVYKNHVSRYRELPIRIAEELGM 360

40 Query: 361 MHRYEKSGALTGLQVRREMTLNDHIFVTPEQIKDEFKALNLIAEIEYEDFNLTDIRFRL 420
 MHRYEKSGAL+GLQVRREMTLND HIPVTPEQI++EF +AL LI ++Y DFNLTDIRFRL
 Sbjct: 361 MHRYEKSGALSGLQVRREMTLNDGHIFVTPEQIQEEFQALQLIIDVYADFNLTDIRFRL 420

45 Query: 421 SYRDPEDKHXYDNDDEMWENQAAMLDKEDFGLDYFEAEGEAAFGPKLDIQVKLTALGN 480
 SYRDP D HKYYDNDDEMWENQA+MLK A+D+ G+DYFEAEGEAAFGPKLDIQVKLTALGN
 Sbjct: 421 SYRDPNDTHKXYDNDDEMWENQA+SMLKAALDEMVDYFEAEGEAAFGPKLDIQVKLTALGN 480

50 Query: 481 EETLSTIQLDFLLPERFDLKYIGADGEEHRPIMIHRRGISTMERFTAILIETYKGAFPTW 540
 EETLSTIQLDFLLPERFDLKYIGADGEEHRP+MIHRG ISTMERFTAILIETYKGAFPTW
 Sbjct: 481 EETLSTIQLDFLLPERFDLKYIGADGEEHRPVMIHRRGVISTMERFTAILIETYKGAFPTW 540

55 Query: 541 LAPQVQSVIPISNEAHIDYAEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKIPYQLI 600
 LAP QV+VIPISNEAHIDYAEVA+ L+DRG+RA+VDDRNEKMQYKIRA+QT KIPYQLI
 Sbjct: 541 LAPHQVTVIPISNEAHIDYAEVAKTLRDRGVRADVDDRNEKMQYKIRASQTSKIPYQLI 600

 Query: 601 VGDKEMEEKAVNVRRYGSKATETKSIEEFVESILADIARKSRPD 644
 VGDKEME+K+VNVRRYGSK T T+S+EEFVE+ILADIARKSRPD
 Sbjct: 601 VGDKEMEDKSVNVRRYGSKTTHTESVEEFVENILADIARKSRPD 644

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 372

60 A DNA sequence (GBSx0403) was identified in *S.agalactiae* <SEQ ID 1213> which encodes the amino acid sequence <SEQ ID 1214>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

-473-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA72250 GB:Y11463 ORF5 [Streptococcus pneumoniae]
 Identities = 189/290 (65%), Positives = 234/290 (80%)

Query: 1 MRIGLFTDTYFPQVSGVSTSIIRTLKEGLEKEGHEVYIFTTDRNVKRFEDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLK LEK+GH V+IFTTDD++V R+ED IIR+PSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIRTLKTELEKQGHAVFIFTTDDKDVNRVEDWQIIRIPSPV 60

15 Query: 61 FISFTDRRVVYRGLISAYRIAKDYELDIHTQTTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
 F +F DRR YRG A IAK Y+LDIHTQTTEFSLGLLG +A+ L+IPV+HTYHTQY
 Sbjct: 61 FFAFKDRRFAYRGRFSALEIAKQYQLDIHTQTTEFSLGLLGIWIARELKIPVIHTYHTQY 120

20 Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSLDLGVICPSRIVNLLDGYGVKIPKQVIPTGIP 180
 EDYV YIAKG LI+PSMVKY++R +L D+DGVICPS IV +LL Y VK+ K+VIPTGI
 Sbjct: 121 EDYVHYIAKGMLIRPSMVKYLVRGFLHDVDGVICPSEIVRDLLSDYKVKVEKRVIPGTIE 180

25 Query: 181 VENYRRREDISEETIKNRLTELGLADNDTMLLSLSRVSEKNIQAAILMHLASAVVDENPHVK 240
 + + R +I +E +K LR++LG+ D + LLSLSR+S+EKNIA+L+ + V+ E VK
 Sbjct: 181 LAKFERPEIKQENLKELRSKLGIDGECTLLSLSRISYEKNIQAVLVAFADVLKEEDKVK 240

30 Query: 241 LVIVGDGPYLSDLKELVHSELENSVIFTGMVEHSQVAIYYKACDFFISA 290
 LV+ GDGPYL+DLKE +LE+++SVIFTGM+ S+ A+YKKA DFFISA
 Sbjct: 241 LVVAGDGPYLNLDLKEQAQNLEIQDSVIFTGMIAPSETALYYKAADFFISA 290

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1215> which encodes the amino acid sequence <SEQ ID 1216>. Analysis of this protein sequence reveals the following:

35 Possible site: 17
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1074(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

 Identities = 309/444 (69%), Positives = 370/444 (82%)

45 Query: 1 MRIGLFTDTYFPQVSGVSTSIIRTLKEGLEKEGHEVYIFTTDRNVKRFEDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLKE LEKEGHEVYIFTTDR+VKRFEDPTIIRLPSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIRTLKEELEKEGHEVYIFTTDRDVKRFEDPTIIRLPSVP 60

50 Query: 61 FISFTDRRVVYRGLISAYRIAKDYELDIHTQTTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
 F+SFTDRRVVYRGLIS+Y+IAK Y LDIHTQTTEFSLGLLGK++ KALRIPVVHTYHTQY
 Sbjct: 61 FVSFTDRRVVYRGLISSYKIAKHYNLDIHTQTTEFSLGLLGKMGKALRIPVVHTYHTQY 120

55 Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSLDLGVICPSRIVNLLDGYGVKIPKQVIPTGIP 180
 EDYV YIA GK+I+PSMVK ++R YL DLDGVICPSRIVNLL+GY V IPK+VIPTGIP
 Sbjct: 121 EDYVSYIANGKIIRPSMVKPLLRGLYKDLDDGVICPSRIVNLLLEGYEVITPKRVIPTGIP 180

 Query: 181 VENYRRREDISEETIKNRLTELGLADNDTMLLSLSRVSEKNIQAAILMHLASAVVDENPHVK 240
 +E Y R+DI+ E + NL+ ELG+A ++TMLLSLSR+S+EKNIAI+ + A++ EN +K
 Sbjct: 181 LEKYIRDDITAEVITNLKAEGLIAGDETMLLSLSRISYEKNIQAIIINQMPAILAENAKIK 240

60 Query: 241 LVIVGDGPYLSDLKELVHSELENSVIFTGMVEHSQVAIYYKACDFFISASTSETQGLTY 300
 L+IVG+GPYL DLK L LE++ V FTGMV H +VA+YKACDFFISASTSETQGLTY
 Sbjct: 241 LIIVGNPXLQDLKHLAMQLEVDKHTFTGMVPHDKVALYYKACDFFISASTSETQGLTY 300

 Query: 301 IESLASGRPIIAQSNPYLDDVISDKMFGTLYKKESDLADAILDAIAETPKMTQEAYEQKL 360

-474-

IESLASG PIIA NPYLDDV++DKMFGLTY E+DL DAI+DAI +TP M + +K
 Sbjct: 301 IESLASGTPIAHGNFYLDDVVTDKMFGLTYAETDLTDAIIDAILKTPVMDKRLAKKR 360

Query: 361 YEISAENFSKSVYAFYLDLFLISQKASVKEKVSILTIGNKDSHSTLRFVRKAVYLPKKVFTF 420
 YEISA++F KS+Y FYLD LI++ + +K+SL + + S+L+ V+ A++LPK+
 Sbjct: 361 YEISAQHFGKSIYTFYLDLTIARNSKEAQKLSLYLNHSGKSSSLKLVQGAIHLPKRAAKV 420

Query: 421 TGRASKKVVKAPKRRIRISSIRDFLD 444
 T S KVVKAP + + +I+DFLD
 Sbjct: 421 TAITSVKVVKAPIKLVHAIKDFLD 444

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 373

15 A DNA sequence (GBSx0404) was identified in *Sagalactiae* <SEQ ID 1217> which encodes the amino acid sequence <SEQ ID 1218>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4076(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 117/350 (33%), Positives = 178/350 (50%), Gaps = 29/350 (8%)

30 Query: 1 MKVLLYLEAEYLLKKSGIGRAIKHQEKALQIAGIDYTTNFT----- 41
 M+ L YLEA E L+ G+ A Q AL+ ++ P
 Sbjct: 2 MRALNYLEAAEALR-GGMVTATNQRAALETDDVEVVETPWRAGDPVRSIGSLAAGGSCF 60

Query: 42 DDFDLVHMNTYGIKSWLLMSKAKKTGKKVIMHGHSTEEFRNSFIGSNLVSPLFKWYLCR 101
 FD+ H N G S + A++T ++H H T EDF SF GS+ ++P + YL
 35 Sbjct: 61 TAFDVAHCNVLVGPGSVAVARHARRTDTPVLVHAHLTREDFAQSFRGSSTIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLRSYQSEKKESAFRHYFHLKDD 161
 FY +AD ++ P++Y+K +++AY + PI LSNG+DL Q E + R F L D
 40 Sbjct: 121 FYSQADLVLCPSYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL--DG 178

Query: 162 KVMVGAGLYFMRKGIDQFVEVAAKMPDIRFIWFGETNKWVIPRKVRQIVTKQHPSNVTF 221
 VV G F RKG+ F E+ AK D F WFG ++ + P+NVTF
 45 Sbjct: 179 TVVYAVGEVFERKGLTMFCCL-AKATDHEFAWFGPYDEGPQAGAAATRKWVADPPANVTFT 237

Query: 222 GYIKGDVYEGAMSASDAFFFPRETEGIVVLEALASHQHVLRDIPVYHGWVTE-DSVE 280
 GY++ A A D + FP++ E +GI VLEA+A + VVLRDIPV+ + T+ +
 Sbjct: 238 GYMEDK--RAAFGAGDIYLFPAKVENQGIADVLEAMACGKPVVLRDIPVFREFFTDGEDCL 295

50 Query: 281 LATDVDGFVEKLDKVLGSKSDKIKEGYH---VAESRSIERIAHELASVYQ 327
 + + + F + +D++ + + G + AES S++RI ELAS+Y+
 Sbjct: 296 MCSTFEAFRDAIDRLADDPELRTRLGENARETAESHSLDRIGEELASIYE 345

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1219> which encodes the amino acid sequence <SEQ ID 1220>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-475-

bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 236/332 (71%), Positives = 276/332 (83%)

Query: 1 MKVLLYLEAEEYLKKSGIGRAIKHQEKALQIAGIDYTTNPTDDFDLVHMNTYGIRSWLLM 60
 MKVLLYLEAE YL+KSGIGRAIKHQ KAL + G +TNP + +DLVH+NTYG++SWLLM
 10 Sbjct: 1 MKVLLYLEAENVLRKSGIGRAIKHQAKALSLVGQHFTTNPRETYDLVHLNTYGLKSWLLM 60

Query: 61 SKAKKTGKKVIMHGHSTEEDEFNSFIGSNLVSPLFKWYLCRFYQKADAIITPTDYSKQLI 120
 KA+K GKVMHGHSTEEDEFNSFI SNL+SP FK YLC FY KADAIITPT YSK LI
 15 Sbjct: 61 IKAQKAGKKVIMHGHSTEEDEFNSFIFSNLLSPWFKYLCHFYKADAIITPTLYSKSLI 120

Query: 121 KAYGIKKPIFVLSNGIDL SRYQRSEKESAFRHYFHL SKDDKVVMGAGLYFMRKGIDQFV 180
 ++YG+K PIF +SNGIDL +Y KKE+AFR YF + + +KVVMGAGL+F+RKGD FV
 20 Sbjct: 121 ESYGVKSPIFAVSNGIDLEQYGADPKKEA AFRYFDIKBGEKVVMGAGLFFLRKGIDDFV 180

Query: 181 EVAAKMPDIRFIWFGETNKWVIPRQVRQIVTKQHPSNVTFAGYIKGDVYEGAMSADAFF 240
 +VA MPD+RFIWFGETNKWVIP +VRQ+V HP N+ F GYIKGDVYEGAM+ +DAFF
 25 Sbjct: 181 KVAQAMPDVRFIWFGETNKWVIPAQVRQMVNGNHPKNLIFPGYIKGDVYEGAMTGADAFF 240

Query: 241 FPSREETEGIVLEALASHQHVLRDIPVYHGWVTEDSVELATDVGDFVEKLDKVLSGKS 300
 FPSREETEGIVLEALAS QH+VLRDIPVY+GWV + S ELATD+ GF+E L KV SG S
 25 Sbjct: 241 FPSREETEGIVLEALASRQHLVLRDIPVYVYQVQSSAELATDIPGFIEALKKVFSGAS 300

Query: 301 DKIKEGYHVAESRSIERIAHELASVYQKVMEL 332
 +K++ GY VA+SR +E + H L VY+KVMEL
 30 Sbjct: 301 NKVEAGYKVAQSRRLTVGHALVDVYKVMEL 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 374

35 A DNA sequence (GBSx0405) was identified in *S.agalactiae* <SEQ ID 1221> which encodes the amino acid sequence <SEQ ID 1222>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5487(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]
 Identities = 308/483 (63%), Positives = 378/483 (77%)

50 Query: 1 MTNELIMQAFEWYLPDGNHWWKLEESISDLKKLGISKIWLPPAFKGTSSDDVGYGVYDL 60
 MTNE +MQ FEWYLP+DG HW+ L E S LK +GISK+W+PPAFKGT S+DVGYGVYDL
 Sbjct: 1 MTNETMMQYFEWYLPNDGKHWHQHLAEDASHLKNIGISKVWMPAFKGTGSNDVGYGVYDL 60

Query: 61 FDLGEFDQNGTIRTKYGRKEEYLLKIKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVN 120
 +DLGEF+QNGT+RTKYG +E+YL + +LK I P +DIVLNHKANGD KE+FOV+KVN
 55 Sbjct: 61 YDLGEFNQNGTVRTKYGSREDYLNNAVNALKEQEIMPISDIVLNHKANGDAKERFQVVKVN 120

Query: 121 PENRQEALSEPVEIEGWTGFDPPGRQGEYNDFKWHWHYFTGLDYDAKNNETDIFMIVGDN 180
 P NRQE +SEPVEIEGWT F+FPGRQ Y+DFKWHWHYFTG+DYDA +NE I+MI+GDN
 60 Sbjct: 121 PSNRQEKISEPVEIEGWTQFNFPGRQDNYSDFKWHWHYFTGVDYDALHNENGIYMLGDN 180

Query: 181 KGWADDDLIDDENGNFDYLMYNDIDFKHPEVIKNLQDWAKWFIETTTGIEGFRDVAVKHID 240

-476-

KGWA + ID ENGN+DYLMY+DIDFKHPEV ++L+DW WF+ET+G+ GFRLLDA+KHID
 Sbjct: 181 KCWASQENIDQENGNYDYLMYDDIDFKHPEVQEHLRDWVAFLETSVGGFRLDAIKHID 240

5 Query: 241 SYFIQTFINDIRTKIKPDLEVFGEYWKSDQTSMDYLEATQFQFSLVDVTLHMNFFDASH 300
 F+ FI IR +K DL VFGEYWK + DYL + QF L+DV LHM+ F+A
 Sbjct: 241 KITFMAQFIRYIREHLKADLYVFGEYWKDSHFDTITDYLHSDVLQFDLIDVMLHMSLFEGQ 300

10 Query: 301 QNRDFDMRTIFDDSLVIDNPYAVTFVENHDTQSGQALERSVEDWFKPLAYGLILLRQQG 360
 + DFD+ TI DDSL+ +P++AVTFV+NED+Q GQALES V +WFKPLAYGLILLRQ+G
 Sbjct: 301 KGSDFDLSTILDDSLMKSHPDFAVTFVDNHSQRGQALESTVAEWFKPLAYGLILLRQEG 360

15 Query: 361 TPCLFYGDYYGIQGEFGQPSFKEVIDKMAELRQNYVFGKQVDYFTHSNCIGWTCLGDEEH 420
 PC+FYGDYYGI GEF Q SF+ V+DK+ +RQ +V+G + T NCIGWTCLGDEEH
 Sbjct: 361 IPCVFGDYGGISGEFAQESFQTVLKLILYIRQYHVGSKKIIITMPNCIGWTCLGDEEH 420

Query: 421 NSCLAVVLTNGDQGWKHMEVGEIYAGKTFVDYLGNCQEVEVIGDDGWGDFLIVESASISAW 480
 +AV+++NG+ K M +GE K FVDYL NC +EV++ D GWGDF V+ AS+SAW
 Sbjct: 421 PDGVAVILISNGEANCRRMMGEFNRNKVFVDYLNNCTEEVLDDQGWGDFPVQEASLSAW 480

20 Query: 481 VPK 483
 V K
 Sbjct: 481 VNK 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1223> which encodes the amino acid
 25 sequence <SEQ ID 1224>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:AAB00845 GB:M57692 alpha-cyclodextrin glycosyltransferase
 [Thermoanaerobacterium thermosulfurigenes]
 Identities = 356/710 (50%), Positives = 468/710 (65%), Gaps = 16/710 (2%)

40 Query: 7 KITKLLTKSAVLLGLISFPLT--VSAADNASVTNKADFSTDTIYQIVTDRFNDGNTSNN 64
 KT+KL+ + L L+ F LT + AA + +V+N ++STD IYQIVTDRF DGNTSNN
 Sbjct: 3 KITKLLTKSAVLLGLISFPLT--VSAADNASVTNKADFSTDTIYQIVTDRFNDGNTSNN 61

45 Query: 65 KTDVFDKN--DLKKYHGGDWQGIIAKIKDGYLTDMGISAIWISSPVENIDSIDPSN---G 119
 D++D LKKY GGDWQGI KI DGYLT MG++AIWIS PVENI ++ P + G
 Sbjct: 62 TGDLYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGVTAIWISQPVENIYAVLPDSTFGG 121

Query: 120 SAAYHGYWAKDFFKINQHFGEADFOQLVKVAHQHHIKVVIDFAPNHTSTAEKEGTTFKE 179
 S +YHGYWA+DF +TN +FG+ DFG L+ AH H+IKV+IDFAPNHTS A + T+ E
 50 Sbjct: 122 STSYHGYWARDFKRTINPYFGSFTDFQNLINTAHANIKVVIDFAPNHTSPASETDPYAE 181

Query: 180 DGALYKNGKLVGKFSDDKDKIFNHESWTDFTSYENSIYHSMYGLADLNNINPKVDQYMKE 239
 +G LY NG L+G +++D + F+H TDFS+YE+ IY +++ LADLN N +D Y+K
 55 Sbjct: 182 NGRLYDNGTLLGGYTNDTNGYFHHYGGTDFSSYEDGIYRNLFDLADLNQQNSTIDSYLKS 241

Query: 240 AIDKWLDLGVDGIRVDVAVKHMSSQGWQKNWLSHIYEKHNVFVGEWFSGHTDDDDYDMITFA 299
 AI WLD+G+DGIR+DAVKHM GWQKN++ I VF FGEWF G + D + T FA
 Sbjct: 242 AIKVLDMGIDGIRLDVAVKHMFPGWQKNFMDLSLSYRPVFTFGEWFLGTNEIDVNNYFA 301

60 Query: 300 NNSGMGLLDFRANAIQLYTGFSFTMRDFYKVLNDRDQVTNEVTDQVTFIDNHDMERF 359
 N SGM LLDLFR+ +RQ++ +T TM ++++ N + D VTFIDNHD+RF
 Sbjct: 302 NESGMGLLDFRFSQKVRQVFRD--NTDTMYGLDSMIQSTASDYNFINDMVTFIDNHDMDRF 360

65 Query: 360 ATKVANNQTAVNQAYALLLTSRGVPNIYYGTEQYATGDKDPNNRGDMPSFNKESQAYKVI 419
 + V QA A LISRGPV IYYGTEQY TG+ DP NR M SFN + AY VI

-477-

Sbjct: 361 YN--GGSTRPVEQALAFITLTSRGVPAIYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVI 418

Query: 420 SKLAPLRKQNALAYGTTEQRWISDHVLVFERKFGNHVALVAINRDQINGYITITNAKTAL 479
KLAPLRK N A+AYGTT+QRWI++ V ++ERKFGN+VALVAINR+ + Y IT TAL

5 Sbjct: 419 KKLAPLRKSNPAIAYGTTQQRWINNDVYIERKFGNNVALVAINRNLSTSYNITGLYTAL 478

Query: 480 PQNSYKDKLEGLLGGQRLIVGADGTISSFELGAGQVAVWVTEGEDKTPQLGDVDASVGIA 539
P +Y D L GLL G + V +DG+++ F L AG+VAVW Y +P +G V ++ A

10 Sbjct: 479 BAGTYTDVLGGLLNGNSISVASDGSVTPFTLSAGEVAVWQYVSSSNPLIGHVGPTMTKA 538

Query: 540 GNKITISGQGFNGSKQVTFGEISAEILSWSDTLITLKVPTVPANYVNISVTTADKQTSN 599
G ITI G+GFG + GQV FG + I+SW DT + +KVP+V YNIS+ T+ TSN

15 Sbjct: 539 GQTITIDGRGFGTTSQVLFGSTAGTIVSWDDTEVKVKVSVTPGKYNISLKTSSGATSN 598

Query: 600 SYQAFEVLTDKQIPVRLINDFKTVPGEQLYLMGDFEMGANDAKNAVGLFNNTQTIK 659
+Y +LT QI VR ++N+ TV GE +YL G+V E+G D A+GP+FN Q + +

20 Sbjct: 599 TYNNINILTNQICVRFVNNASTVYGENVYLTGNVAELGNWDTSKAIGPMFN--QVVYQ 656

Query: 660 YPNWFFDTHLPINKEIAVKLVKKDSIGNVLWT--SPETYSIKTGHEAQT 707
YP W++D +P I K +KK+ + W S TY++ + I

Sbjct: 657 YPTWYDVSVPACTTIQFKFIKNG-NTITWEGGSNHTYTVPSSTGTVI 705

An alignment of the GAS and GBS proteins is shown below:

Identities = 112/509 (22%), Positives = 193/509 (37%), Gaps = 103/509 (20%)

25 Query: 18 GNHWKLEESISD--LKKLGISKIWLPPAFKGTSSDDV-----GYGVYDLFDLGEFD 67
G W+ + I D L +GIS IW+ + S D GY D F +
Sbjct: 79 GGDWQGIIAKIKDGYLTDMGISAIWISSPFVENIDSIDPSNGSAAYHGYWAKDFFKTNQH- 137

30 Query: 68 QNGTIRTKYGRKEEYKLKLSLKANGIKPFADIVLNHKANGDHKEKFQVIKVPENRQEA 127
+G + ++ +L+K + IK D NH + + +
Sbjct: 138 -----FGTEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEKE----- 173

35 Query: 128 LSEPYEIEGWTGFDFFGRQGEYNDFKWHWHYHFTGLDYDAKNNETDIEMIVGDNKGWADDD 187
G F Y + K G D K+ + +++ W D
Sbjct: 174 -----GTTFKEDGALYKNGK-----LVGKFSDDKDK-----IFNHESWTDPS 210

40 Query: 188 LIDDE--NGNFDYLMYNDIDFKHPEVIKNLQDWAKWFIETTGIEGFRLDVAVKHIDSYFIQ 245
++ + + N+I+ K + +K D KW G++G R+DAVKH+ + +
Sbjct: 211 TYENSIYHSMYGLADLNNINPKVDQYMKEAID--KWL--DLGVDGIRVDVAVKHMSQGWQK 266

45 Query: 246 TFINDIRTKIKPDLEVFGEYWKSDQTSMDYLEATQFQFSLVDVTLHMNFDDASHQ-NRD 304
+++ I K ++ VFE W S T D + T F + L F +A Q
Sbjct: 267 NWLSHIYE--KHNVFVFE-WFSGHTD--DDYDMTTFANNSGMLDPRFANAIRQLYTG 321

50 Query: 305 FDMRTIFDDSLVIDNPEYA-----VTFVENHDTQSGQALESERVEDWFKPLAYGLILLR 357
F T+ D V++N + VTF++NHD + + + AY L LL
Sbjct: 322 FSTFTMRDFYKVLNDRDQVTNEVTDQVTFIDNHDMERFATKVANNQTAVNQ-AYAL-LLT 379

55 Query: 358 QQGTPCLFYGDYGIQGE-----FGQPSFK-----EVIDKMAELR---QNYVFGKQVD 402
+G P ++YG G+ PSF +VI K+A LR Q +G
Sbjct: 380 SRGVPNIYGTQYATGDKDPNNGDMPNKFESQAYKVISKLAPLRKQNALAYGTTEQ 439

60 Query: 403 YFTHSNCIGWTCIGDEEHNSCLAVLTNGDQ--GWKHMEVGEIYAGKTFVDYLGNC--EQ 458
+ + + + + +A+V N DQ G+ ++ D L Q
Sbjct: 440 RWISDHVL---VFERKFGNHVALVAINRDQINGYITITNAKTALPQNSYKDKLEGLLGGQ 495

Query: 459 EVVIGDDGW-GDFLVESASISAWVEKIEE 486
E+++G DG F + + ++ W + E+

Sbjct: 496 ELIVGADGTISSFELGAGQVAVWVTEGED 524

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-478-

Example 375

A DNA sequence (GBSx0406) was identified in *S.agalactiae* <SEQ ID 1225> which encodes the amino acid sequence <SEQ ID 1226>. This protein is predicted to be catabolite control protein A. Analysis of this protein sequence reveals the following:

```

5   Possible site: 29
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2154(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9707> which encodes amino acid sequence <SEQ ID 9708> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAA88121 GB:AB028599 catabolite control protein A [Streptococcus
    bovis] (ver 3)
    Identities = 304/332 (91%), Positives = 320/332 (95%)

20 Query: 1  MNTDDTITIIDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA 60
    MNTDDTITIIDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA
    Sbjct: 1  MNTDDTITIIDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA 60

25 Query: 61  SKKTTTVGVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDKENVVNTLFAKQ 120
    SKKTTTVGVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDKENVVNTLFAKQ
    Sbjct: 61  SKKTTTVGVVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDKENVVNTLFAKQ 120

30 Query: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVLAGTVDLEHQLPSVNIDYKAAAVDVIDLAGNH 180
    VDGIIFMGHHLTEKIRAEFSRSRTP+VLAGTVDLEHQLPSVNIDYKAA DV+DILA N+
    Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVVLVAGTVDLEHQLPSVNIDYKAAAVDVVDILAKNN 180

35 Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNLNFKEGLVFEANYRYAEGFALAQRVINAG 240
    KDIAFVSGPLIDDINGKVRLAGYKEGL+KN L+FKEGLVFEANY Y +G+ LAQRV+N+G
    Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLEKNLSPFKEGLVFEANYNYKDGIELAQRVMNSG 240

40 Query: 241 ATAAYVAEDELAAAGLLNGLFEAGKRVPEDFEITTSNDSPIAQYTRPNLTSISQPVYDLGA 300
    ATAAYVAEDELAAAGLLNGLF AGK+VPEDFEI+TSNDSPI YTRPNL+SIGQPVYDLGA
    Sbjct: 241 ATAAYVAEDELAAAGLLNGLFAAGKKVPEDFEILTSDSPITSYTRPNLSSISQPVYDLGA 300

40 Query: 301 VSMRMLTKIMHKEELEKEEIVLNHGIVKRGTT 332
    VSMRMLTKIM+KEELEKEE+LNHG+ RGGT
    Sbjct: 301 VSMRMLTKIMNKEELEKEEILNHLGLKLRGTT 332

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1227> which encodes the amino acid sequence <SEQ ID 1228>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.2154(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55      Identities = 307/332 (92%), Positives = 320/332 (95%)

Query: 1  MNTDDTITIIDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA 60
      MNTDD +TIIYDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA
Sbjct: 1  MNTDDPLTIIYDVAREAGVSMATVSRVVGNGKNVKENTRKKVLEVIDRLDYRPNVARGLA 60

```


-479-

Query: 61 SKKTTTVGVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120
 Sbjct: 61 SKKTTTVGVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDDKEVNVVNTLFAKQ 120
 Query: 121 VDGIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDLHQLPSVNIDYKAAAVDVIDILAGNH 180
 Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTP+VLACTVDL+HQLPSVNIDY+AA +V+DILA NH 180
 Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNGLNFKEGLVFEANYRYAEGFALAQRVINAG 240
 Sbjct: 181 K IAFVSGPLIDDINGKVRLAGYKEGLK N L+FKEGLVFEANY Y EGF LAQRVIN+G 240
 Query: 241 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEITSDNSPIAQYTRPNLTSISQPVVDLGA 300
 Sbjct: 241 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEITSDNSP+ QYTRPNL+SIGPVVDLGA 300
 Query: 301 VSMRMLTKIMHKBELEKEIVLNHGIVKRGTT 332
 Sbjct: 301 VSMRMLTKIM+KEELEKEEI+LNHGI KRGTT 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 376

A DNA sequence (GBSx0407) was identified in *Sagalactiae* <SEQ ID 1229> which encodes the amino acid sequence <SEQ ID 1230>. This protein is predicted to be PepQ (pepQ-2). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1118(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 257/359 (71%), Positives = 304/359 (84%)

Query: 1 MSKLNRIHHLHSVQAEIAVSDPVTNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
 MSKL +I L E AV SDPV++NYLTGF+ DPHER MFLF++ D++ +LF+P L+
 Sbjct: 1 MSKLAQIVQKLKKQIEAAVLSDPVSINYLTFYSDPHERLMFLFLFADQETLLFLPELD 60
 Query: 61 VSRAKQSVFPFVFGYIDSENWPQKIASNLPSFSVSKVLAEFDNLVTKFQGLQTVFDGHF 120
 RAK + V GY+D ENP +KI + LP + SK+ EFDNLNVTKF+GL+T+F G F
 Sbjct: 61 ALRAKSILDISVTGYLDFENPLEKIKTLLEKTNYSKIALEFDNLNVTKFQGLTIFSGQF 120
 Query: 121 ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFNDNISLNTETDIIAQIEFEMKKQ 180
 NLTP I MRLIKS DEI+K+L+AGE ADKAVQ+GFD+ISLN TETDIIAQIEFEMKK
 Sbjct: 121 TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180
 Query: 181 GINKMSFDTMVLTGNNANPHGIPGINKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240
 G++KMSF+TMVLTG+NAANPHG+P ++KIENN LLLFDLGV+ GY SDMTRTVAVG+PD
 Sbjct: 181 GVDKMSFETMVLTGNSAANPHGLPASHKIENNHLFFFDLGVSTGYVSDMTRTVAVGQPD 240
 Query: 241 QFKKDIYHLCLEAHQAIDFIKPGVLASEVDAAARNVIEKAGYGYFNNHRLGHGLGMDVH 300
 QFKKDIY++CLEA A+DFIKPGV A++VDAAAR+VIEKAGYG YFNNHRLGHG+GM +H
 Sbjct: 241 QFKKDIYNICLEAQLTALDFIKPGVSAAQVDAARSVIEKAGYGDYFNNHRLGHGIGMGLH 300
 Query: 301 EFPSIMAGNDMEIQEGMCFSEPGIYIPDKVGVRIEDCGYVTKTGFEVFTKTPKELLYF 359
 EFPSIMAGNDM ++EGMCFSEPGIYIP+KVGVRIEDCG+VTK GFEVFT+TPKELLYF

-480-

Sbjct: 301 EFPSIMAGNDMLLEEGMCFSEVEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1231> which encodes the amino acid sequence <SEQ ID 1232>. Analysis of this protein sequence reveals the following:

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.90 Transmembrane 42 - 58 (42 - 59)
 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 264/359 (73%), Positives = 304/359 (84%)
 Query: 1 MTKLDQIRLYLDQKGAEIAIFSDPVTINYLTFGFFCDPHERQLFLFVYHDLAPVLFVPALE 60
 M+KL QI L ++G E A+ SDPV+INYLTFG+ D PHER +FLF++ D +LF+P L+
 20 Sbjct: 1 MSKLAQIVQKLKKQGIEAAVLSDPVSINYLTFGYSDPHERLMFLFLFADQETLLFLPELD 60
 Query: 61 VARASQAISFPVFGYVDSENPEWIKAVLPNTAAKTIYAEFDHLNVNKFHGLQTIIFSGQF 120
 RA + V GY+D ENP BKIK +LP T I EFD+LNV KF GL+TIFSGQF
 25 Sbjct: 61 ALRAKSILDISVTGYLDFENPLEKIKITLLPKTNYSKIALFEFDNLNVTKFKGLETIIFSGQF 120
 Query: 121 NNLTPYVQGMRLVKSADETINKMMIAGQFADKAVQVGFNDISLDATETDVIAQIEFEMKKQ 180
 NLTP + MRL+KSADEI K++IAG+ ADKAVQ+GFD+ISL+ATETD+IAQIEFEMKK
 30 Sbjct: 121 TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180
 Query: 181 GIHKMSFDTMVLGTGNNANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240
 G+ KMSF+TMVLGT+NAANPHG+P ++ IENN LLLFDLGV+ GY SDMTRTVAVGQPD
 35 Sbjct: 181 GVDKMSFETMVLGTGSNAANPHGLPASHKIENNHLALLFDLGVESTGYVSDMTRTVAVGQPD 240
 Query: 241 QFKIDIYNLCLEAQLAAIDFIKPGVTAQVDAARQVIEKAGYGEYFNHRLGHGIGMDVH 300
 QFK DIYN+CLEAQL A+DFIKPGV+AAQVDAAR VIEKAGY+YFNHRLGHGIGM +H
 40 Sbjct: 241 QFKKDIYNLCLEAQLTALDFIKPGVSAQVDAARSVIEKAGYGDYFNHRLGHGIGMGLH 300
 Query: 301 EFPSIMAGNDMLLEEGMCFSEVEPGIYIPEKVGVRIEDCGHVTKNGFEVFTHTPKELLYF 359
 EFPSIMAGND++LEEGMCFSEVEPGIYIP KVGVRIEDCGHVTKNGFEVFT TPKELLYF
 45 Sbjct: 301 EFPSIMAGNDMLLEEGMCFSEVEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

An alignment of the GAS and GBS proteins is shown below:

Identities = 288/361 (79%), Positives = 325/361 (89%)
 45 Query: 1 MSKLNIRIRHHLHSVQAEIAVFSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
 M+KL++IR +L AELA+FSDPVT+NYLTGFFCDPHERQ+FLFVY D P+LFVPALE
 Sbjct: 1 MTKLDQIRLYLDQKGAEIAIFSDPVTINYLTFGFFCDPHERQLFLFVYHDLAPVLFVPALE 60
 Query: 61 VSRAKQSVFPVFGYIDSENWQKIASNLPFSVSKVLAEFDNLNVTKFQGLQTVFDGHF 120
 V+RA Q++ FPVFGY+DSENW+KI + LP+ + + AEFD+LNV KF GLQT+F G F
 50 Sbjct: 61 VARASQAISFPVFGYVDSENPEWIKAVLPNTAAKTIYAEFDHLNVNKFHGLQTIIFSGQF 120
 Query: 121 ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFNDISLNTTETDIIAQIEFEMKKQ 180
 NLTPY+Q MRL+KS DEI KM++AG+ADKAVQVGFNDISL+ TETD+IAQIEFEMKKQ
 55 Sbjct: 121 NNLTPYVQGMRLVKSADETINKMMIAGQFADKAVQVGFNDISLDATETDVIAQIEFEMKKQ 180
 Query: 181 GINKMSFDTMVLGTGNNANPHGIPGTNKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240
 GI+KMSFDTMVLGTGNNANPHGIPGTN IENNALLLFDLGVETLGYTSDMTRTVAVG+PD
 60 Sbjct: 181 GIHKMSFDTMVLGTGNNANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240
 Query: 241 QFKKDIYHLCLEAHQAIDFIKPGVLAASEVDAARNVIEKAGYGYFNHRLGHGIGMDVH 300
 QFK DIY+LCLEA AIDFIKPGV A++VDAAR VIEKAGY+YFNHRLGHG+CMDVH
 Sbjct: 241 QFKIDIYNLCLEAQLAAIDFIKPGVTAQVDAARQVIEKAGYGEYFNHRLGHGIGMDVH 300

-481-

Query: 301 EFPSIMAGNDMEIQEGMCFSEVPGIYIPDKVGVRIEDCGYVTKGFEVFTKTPKELLYFEG 361
 EFPSIMAGND+ ++EGMCFSEVPGIYIP KVGVRIEDCG+VTK GFEVFT TPKELLYFEG
 Sbjct: 301 EFPSIMAGNDLVLEEGMCFSEVPGIYIPGKVGVRIEDCGHVTKNGFEVFTHTPKELLYFEG 361

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 377

A DNA sequence (GBSx0408) was identified in *S.agalactiae* <SEQ ID 1233> which encodes the amino acid sequence <SEQ ID 1234>. Analysis of this protein sequence reveals the following:

10 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3629(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 378

A DNA sequence (GBSx0409) was identified in *S.agalactiae* <SEQ ID 1235> which encodes the amino acid sequence <SEQ ID 1236>. This protein is predicted to be beta-hexosamidase A precursor. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3279(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAB11942 GB:Z99104 alternate gene name: yzbA~similar to
 beta-hexosaminidase [Bacillus subtilis]
 Identities = 151/602 (25%), Positives = 268/602 (44%), Gaps = 69/602 (11%)

40 Query: 26 INEMTLDEKIGQLF-----FNMGASRSEFYLTVDLRYHIAAVRYNRGSSSEIYDQNL- 78
 +N M+LDEK+GQ+ + S + LT + D +Y G ++ +N+
 Sbjct: 39 VNRMSLDEKLGQMLMPDFRNWQKEGESSPQALTKMNEVASLVKKYQFGGI-LFAENVK 97

Query: 79 -----ILQTKSKLPLMLIAANTEAGGDGAVTDGTVKGDEIKVAATNDPKYAYEMG 127
 + K+P++++ + E G + +GT + + A AY+ G
 45 Sbjct: 98 TTKQIVQLTDDYQKASPKIPLMLSIDQEGGIVTRLGEGTNFPGNMALGAARSINAYQTG 157

Query: 128 RIAGMEASAVGCNASFSPIVDLTRNWRNPPIASRNWGANVDQIISLSKEYMKGIMQYNIV 187
 I G E SA+G N FSP+VD+ N NP+I R++ +N + L MKG+ + +I
 Sbjct: 158 SIIGKELSALGINTDFSPVVDINNPNPNVIGVRSFSSNRELTSRLGLYTMKGLQRQDIA 217

50 Query: 188 PFAKHFPDGDIDERDHHLSFASNPMSEWMTSTFGRIYGEADAGLPVGMAGHIHLENVE 247
 KHFPD G + D H +E + + DAG VM H+ P +
 Sbjct: 218 SALKHFPDGHCDTDVDSHYGLPLVSHGQERLREVELYPFQKAIDAGADMVMTAHVQFPAPD 277

-482-

Query: 248 KEMHPER--DLDDMLPASLNKTLLEDELLRGEELGYNGAIVTDASHMVGMTASMARDDLPT 305
 + + D ++PA+L+K ++ LLR E+G+NG IVTDA +M + + + +
 Sbjct: 278 DTTYKSKLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAADHFGQEEAVVM 337

5 Query: 306 AIEAGCDLFLF---FNDPDED-----IQWMKEGYEKGILTEERLHDALRRTLGLKAKLG 356
 A++AG D+ L E+ IQ +KE + G + E+++++ R + LK K G
 Sbjct: 338 AVKAGVDIALMPASVTSLEEQKFARVIQALKEAVKNGDIPBQQINNSVERIISLKIKRG 397

10 Query: 357 LHNVEGRQELFMPK-DKAMALINTLESQKIADEVADKAVTLVKDKQKDIFFVNPERYRH 415
 + Y R + K KA ++ + + K ++A+KAVT++K++Q + P P++
 Sbjct: 398 M--YPARNSDSTKEKIAKAKKIVGSKQHLKAEKKLAEKAVTVLKNEQHTL-PFKPKKGSR 454

15 Query: 416 ILLVNVVEGYKGGFGAMIAGNKQRASDYMKE-----LLEARGHEVTWESTEERIMKLPQ 469
 IL+V + A +Q D +K L V+++ E+ +K
 Sbjct: 455 ILIV-----APYEEQTASIEQTIHDLIKRKKIKPVSLSKMNFASQVFKTEHEKQVK--- 505

20 Query: 470 EERAAAIANVYAQK-QPIANLTEHYDLIINLVDVNAGGTTQRIIWPAAGTDPDQPFYVHE 528
 E I Y K P+ N D +I+ D + + ++P A + H
 Sbjct: 506 -EADYIITGSYVVKNDPVVN-----DGVID--DTISDSSKWATVFPRA---VMKAALQHN 554

25 Query: 529 IPSIVISVQHAFALADMPQVGTVINAYD-----GLPSTISAVVAKLAGSEFTGVSP 580
 P +++S+++ + A+ + I Y L I A V + G+++ G P
 Sbjct: 555 KPFFVMSLRNPYDAANFEEAKALIAVVGFKGYANGRYLQPNIPAGVMAIFGQAKPKGTLP 614

Query: 581 VD 582
 VD
 Sbjct: 615 VD 616

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 A related GBS gene <SEQ ID 8565> and protein <SEQ ID 8566> were also identified. Analysis of this protein sequence reveals the following homology to a lipoprotein, with homology with the following sequences in the databases:

29.5/52.3% over 422aa

Bacillus subtilis

35 EGAD|20114| hypothetical 70.6 kd protein in feua 5'region precursor Insert characterized
 SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW INTERGENIC REGION
 PRECURSOR (ORF1). Insert
 characterized
 GP|1944006|dbj|BAA19499.1||AB002150 Ybbd Insert characterized

40 GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature protein may
 be membrane-anchored and start at Cys-17. 17.5% identity
 over 354-aa overlap with Candida pelliculosa beta-glucosidase.; putative Insert
 characterized
 GP|2632433|emb Insert characterized

45 ORF00431(367 - 1557 of 2388)
 EGAD|20114|BS0166(36 - 458 of 642) hypothetical 70.6 kd protein in feua 5'region precursor
 {Bacillus subtilis} SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW
 INTERGENIC REGION PRECURSOR (ORF1). GP|1944006|dbj|BAA19499.1||AB002150 Ybbd {Bacillus
 50 subtilis} GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature
 protein may be membrane-anchored and start at Cys-17. 17.5% identity over 354-aa overlap
 with Candida pelliculosa beta-glucosidase.; putative {Bacillus subtilis} GP|2632433|emb
 %Match = 9.6
 %Identity = 29.5 %Similarity = 52.2
 55 Matches = 119 Mismatches = 183 Conservative Sub.s = 92

114 144 174 204 234 264 294 324
 LMVGDSLGLDAAEQNGIAFYFVLVGKEVKSWEILREDIGEAFAGQFEQQRQKESINTFWANLDN**KG*AMTHLVDLT

60 MRPVFPLILSAVLFLSCFFGA
 10 20

354 384 414 426 456 486 528
 KKPFLNLQEAIEWIEKTINEMTLDEKIGQLFF-----NMGASRSEEYLTVDLDRYHIAAVRYNRGS-----SSEIYDQ

-483-

```

      :      :      : : | : | | : | : :      | : | : |      : |      : :      |
RQTEASASKRAIDANQIVNRMSLDEKLGQMLMPDFRNWQKEGESSQALTKMNDENVASLVKKYQFGGIILFAENVKTKTQ
      40      50      60      70      80      90      100

5   543      573      603      633      663      693      723      753
NLIL-----QTKSKLPMLIAANTEAGGDGAVTDGTVKVGDEIKVAATNDPKYAYEMGRIAGMEASAVGCNASFSPIVDLTR
      : |      : : | : : : : | |      : : |      : | : | | | : | | : | : | : |
TVQLTDDYQKASPKIPLMLSIDQEGGIVTRLGEGTNFPGNMALGAARSINAYQTGSIIGKELSGALGINTDFSPVVDINN
      120      130      140      150      160      170      180

10  783      813      843      873      903      933      963      993
NWRNPILASRNWGANVDQIISLSKEYMKGIMQYNIVPFAKHFPDGDIDERDHHLSFASNPMMSKEEWMSTFGRIYGELEDA
      | | : | : : : | :      | | : : : | | | | : | : : : |
NEDNPVIGVRSFSSNRELT SRLGLYTMKGLQKDIAALKHFPFGHGDITDVSHYGLPLVSHGQERLREVELYPFQKAIDA
      200      210      220      230      240      250      260

15  1023      1053      1080      1107      1137      1167      1197      1227
GLPGVMAGHIHLPNVEKEMHPER-DLDDML-PASLNKTLDELRLGELGYNGAIVTDASHMVGMTASMARDDLPTAIEA
      | | | : : : : : : : | : | : : : | | | : : : | : : : : : | : : : : : |
GADMVMTAHVQFPADFDTTKSKLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAIADHFGQBEAVVMVAKA
      280      290      300      310      320      330      340

20  1290      1320      1350      1380      1410      1437
GCDLFLF-----FNDPDE---DIQWMKEGYEKGILTEERLHDALRRTLGLKAKLGLHNYEGRRQELFMPK-DKAMALIN
      | : |      : : :      | | : | : : : : : | : | | : | : | : : : | : : :
GVDIALMPASVTSLEKEQKFARVIALKRAVNGDIPEQQINNSVERIISLKI KRGM--YPARNSDSTKEKIAKAKKIVG
      360      370      380      390      400      410

25  1467      1497      1527      1557      1587      1617      1647      1677
TLESQKIADDEVADKAVTLVKDKQKDIFFVNPERYRHILLVNVVEGYKGGFGAMIAGNKQRASDYMKELLEARGHEVTWES
      : : | : : : | | : : : : : | : : | : : : : : : : : : : : : : : : : : :
SKQHLKAEKKLAEKAVTVLKNEQ-HTLPFKPKKGSRLIVAPYEEQTASIEQTIHDLIKRKKIKPVSLSKMNFASQVFKT
      430      440      450      460      470      480      490

```

35 SEQ ID 1236 (GBS50) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 8; MW 69.2kDa).

GBS50-His was purified as shown in Figure 192, lane 5.

The GBS50-His fusion product was purified (Figure 192, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 264), which confirmed that the protein is immunoaccessible on GBS bacteria.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 379

A DNA sequence (GBSx0410) was identified in *S. agalactiae* <SEQ ID 1237> which encodes the amino acid sequence <SEQ ID 1238>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 380

A DNA sequence (GBSx0411) was identified in *S.agalactiae* <SEQ ID 1239> which encodes the amino acid sequence <SEQ ID 1240>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9705> which encodes amino acid sequence <SEQ ID 9706> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC21726 GB:U32690 oxidoreductase [Haemophilus influenzae Rd]
Identities = 197/271 (72%), Positives = 229/271 (83%)

Query: 26 NKVVVITGAGGVLGCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIAKAYKANVL 85
NK+++ITGAGGVLC ++AK+ A A +ALLDLN EAA A EI + GG AKAYK NVL
Sbjct: 15 NKLIIITGAGGVLCSEFLAKQLAYTKANIALLDLNFEEADKVAKEINQSGGKAKAYKTNVL 74

Query: 86 SKENLEEVEHQAVLEDLGPTDILVNGAGGNNPKATTDNEFHEDLPSETKTFFELDEAGIS 145
EN++EV + D G DIL+NGAGGNNPKATTDNEFH+ DL T+TFF+LD++GI
Sbjct: 75 ELENIKEVRNQIETDFGTCDILINGAGGNNPKATTDNEFHQFDLNETTRTFDLKSGIE 134

Query: 146 FVFNLNLYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 205
FVFNLNLYLG+LLPTQVFA+DM+G++GANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW
Sbjct: 135 FVFNLNLYGSLLEPTQVFAKMDLGKQGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 194

Query: 206 LAVHFSKVGIRCNALAPGFLVTNQNSLLFTEDGQPTARAEEKILNNTPMGRFGEASELIG 265
LAV+FSKVGIRCNALAPGFLV+NQN +LLF +G+PT RA KIL NTPMGRFGE+ EL+G
Sbjct: 195 LAVYFSKVGIRCNALAPGFLVSNQNLALLFDTEGKPTDRANKILNTNTPMGRFGESEELLG 254

Query: 266 GLFFLADEKSSSFVNGVVLPIDGGFAAYSGV 296
L FL DE S+FVNGVVL+DGGF+AYSGV
Sbjct: 255 ALLFLIDENYSAPVNGVVLVPDGGFSAYSGV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1241> which encodes the amino acid sequence <SEQ ID 1242>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0358(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 77/279 (27%), Positives = 125/279 (44%), Gaps = 19/279 (6%)

Query: 18 MSKTITFTNKVVVITGAGGVLGCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIA 77
M + K+ +ITGA + +AK +A+AGA + D+ QE E G A
Sbjct: 1 MENMFSLQGKIALITGASYGIGFELAKAYAQAGATIVFNDIKQELVDKGLAAYRELIGEA 60

Query: 78 KAYKANVLSKENLEEVEHQAVLEDLGPTDILVNGAGGNNPKATTDNEFHEDLPSETKTFF 137

-485-

Y +V + +++++ + +++G DILVN AG
 Sbjct: 61 HGYVCDVTDEAGIQMVMSQIEDEVGAIDILVNNAG-----IIRRTPL 103

Query: 138 ELDEAGISFVFNLYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKA 197
 E+ V +++ + ++ M+ + IINI SM + + AY+ AK
 Sbjct: 104 EMAAEDFRQVIDIDLNAPFIVSKAVLPSMIAGHGKIINICSMSELGRETVSAYAAAKG 163

Query: 198 AISNFTQWLAVHFSKVGIRCNAIAPGFLVTNQNRSLLFTE-DGQPTARAEKILNNTPMGR 256
 + T+ +A F + I+CN I PG++ T Q L + DG + I+ TP R
 Sbjct: 164 GLKMLTKNIASEFGEANIQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAAR 223

Query: 257 FGEASELIGCLFFLADEKSSSFVNGVVLPIIDGGFAAYSG 295
 +G +L G FLA + +S+FVNG +L +DGG AY G
 Sbjct: 224 WGTEDLAGPAVFLASD-ASNFVNGHILYVDGGILAYIG 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 381

A DNA sequence (GBSx0412) was identified in *S.agalactiae* <SEQ ID 1243> which encodes the amino acid sequence <SEQ ID 1244>. This protein is predicted to be D-mannonate dehydrolase (uxuA). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3188(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04425 GB:AP001509 D-mannonate dehydrolase [Bacillus halodurans]
 Identities = 202/343 (58%), Positives = 261/343 (75%)

Query: 1 MEMSFRWYGEDDPVTLENIGQIPTMKGIVTAIYDVPVGEVWSRERIQQLEKEVHAAGLKI 60
 M ++ RW+G D V LE I QIP MKGIV+AIYDV VG VW +E+I LK +E GL +
 Sbjct: 1 MRLTMRWFGPSDKVKLEYIKQIPGMKGIVSAIYDVAVGGVWPKEKILALKNNIERHGLTL 60

Query: 61 SVIESVPVHEDIKLGPRTRDLLIDNYIQTIVKNLAEGIDTICYNFMPVFDWTRTDLAYQY 120
 VIESVPVHEDIKLG+PTRD I+NY QT+++LA GIDT+CYNFMPVFDWTR+ L ++
 Sbjct: 61 DVIESVPVHEDIKLGKPTRDRYIENYKQTLRHLAECGIDTVCYNFMPVFDWTRSQLDFKL 120

Query: 121 PDGSTALIFDETYSKKMDPVNGELSLPGWDASYSKEEMKAIMDAYAEIDEEKLWENLTYF 180
 DGS ALI++E V + +P++GEL LPGWD SY E +K ++ AY +I EE LW++LTYF
 Sbjct: 121 EDGSEALIYEEDVISRTNPLSGELELPGWDSYENESLKGVLQAYKKISEEDLWDHLTYF 180

Query: 181 IKRIIPEAEAVGVKMAIHPDDPPYSIFGLPRIITGLEAIERFVKLYDSKSNGITLCVGSY 240
 ++ I+P A+ VG+KMAIHPDDPP+SIFGLPRI+T +ER + LYDS ++GIT+C GS
 Sbjct: 181 VQAIMFVADEVGIKMAIHPDDPPWSIFGLPRIVTINKANLERLLSLYDSPNHGITMCSGSL 240

Query: 241 ASDPQNDVLEISRRAFELDRVNFVHARNIKLGDGKSFKESAHPSEYGSIDMYEVIKLCHE 300
 ++ ND+ E+ R R++F HARNIK +SF+ESAH SE GS++M ++K H+
 Sbjct: 241 GANEANDLEPMIRHFGGQRIHFAHARNIKRTGPRSFQESAHLEAGSVNMVAMLKAYHD 300

Query: 301 FGFEGAIRPDHGRMIWGETGRPGYGLYDRALGATYVSGLYEAV 343
 GF G +RPDHGRMIWGE GRPGYGLYDRALGATY++G++EAV
 Sbjct: 301 IGFTGELRPDHGRMIWGEKGRPGYGLYDRALGATYLNIGIWEAV 343

No corresponding DNA sequence was identified in *S.pyogenes*.

-486-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 382

A DNA sequence (GBSx0413) was identified in *S.agalactiae* <SEQ ID 1245> which encodes the amino acid sequence <SEQ ID 1246>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2447(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 383

A DNA sequence (GBSx0414) was identified in *S.agalactiae* <SEQ ID 1247> which encodes the amino acid sequence <SEQ ID 1248>. This protein is predicted to be uronate isomerase. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3066(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04424 GB:AP001509 uronate isomerase [Bacillus halodurans]
Identities = 215/465 (46%), Positives = 294/465 (62%), Gaps = 7/465 (1%)

35 Query: 3 ENTETFMLKNQAAIQLYEE-VKRQPIFDYHCHLDPKIDIFEDHIFDNIVDLWLGGDHYKWR 61
 F +E F+L N+ +LY K PI DYHCHL P++I+E+ F+N+ WLGGDHYKWR
 Sbjct: 4 FLSEDFLLMNEYDRELYYTFAKNMPICDYHCHLSPQEIWENKPFENMTKAWLGGDHYKWR 63

40 Query: 62 LMRANGISEAEITGPASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGVNBILTESNA 121
 MR NG+ E ITG A + EK F A+A+T+ + GNP+YHW+ MELK F ++ L E+N
 Sbjct: 64 AMRLNGVREEFITGCAPDKEKFLAWAKTVPKTIGNPLYHWHMELKTYFHFHQPLDETNG 123

45 Query: 122 EEIYHRLNHLKEHKISERRLIADSKVMFIGTIDHPLDTLEWHKKLAADESKTVVAPTF 181
 E ++ N L++ +PR LI S V IGTID P D+L +H+KL AD++F V PTF
 Sbjct: 124 ENVWDACNRLQQAFTPRALIERSNVRAIGTDDPTDLSLLYHQKLQADDTFHVKVIPTF 183

50 Query: 182 RPDEAF-IEHRHFVDFITKLGIDITQKEITDFSTFIAAMEERAIYFAQNGCRASDISFTEI 240
 RPD A IE F D++ KL D+T + + F+ A++ER+ +F ++GCR+SD TE+
 Sbjct: 184 RPDGALKIEQDSFADWVAKLSDVGTGESLDTLDAFLHALKERLTFFDEHGCRSSDHDMTEV 243

50 Query: 241 VFEQTDLELNDLENKVCGEYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHFGALRNN 300
 F + +E E +F K + E K++T + L + Y G+V Q H G +RNN
 Sbjct: 244 PFVEVNEQEAQHIFRKRLANEGLTQVENEKYKTFMLTWLGKEYAARGWVMQWHIGVMRNN 303

Query: 301 HSTIFEKLGADVGVDSLGD-QVALTVNMNRLDLSLVKDSLPMIWNLNPNAYNIAVANT 359

-487-

+S + KLG D G DS+GD Q+A +LLD L K+ +LPK I Y +NP N +A+
 Sbjct: 304 NSRMLHKLGPDTGFDSIGDGQIAHAT--AKLLDLLDKQGALEPKTILYCVNPNANYILASM 361

Query: 360 LANFQANELGVRSYLQFGAGWWFADTKLGMISQMNALAEQGMLANFIGMLTDSRSFSLSYQ 419
 + NF E GVR +QFG+ WWF D GM Q+ LA G+L+NFIFMLTDSRSFSLSY
 Sbjct: 362 IGNF--TESGVRGKVQFGSAWWFNDHIDGMRRQLTDLASVGLLSNFIGMLTDSRSFSLSY 419

Query: 420 RHDYFRRILCTYLGEWIEEGEVPEDYQALGSMADIAVQNAVNYF 464
 RHDYFRRILC +G WI+EG++P D + G + +DI Y N V+YF
 Sbjct: 420 RHDYFRRILCQLIGSWIKEGQLPPDMERWGQIVQDICYNNVVDYF 464

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 384

A DNA sequence (GBSx0415) was identified in *S.galactiae* <SEQ ID 1249> which encodes the amino acid sequence <SEQ ID 1250>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3883(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9703> which encodes amino acid sequence <SEQ ID 9704> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
 aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 93/199 (46%), Positives = 125/199 (62%), Gaps = 6/199 (3%)

Query: 37 KNNYFFAVIRGKSSDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNNKEIIV 96
 K + AV+R S E+A E A GG+ IE+TF+ P+A VIK+LS F K I+
 Sbjct: 8 KKKHIVAVLRANSVEEAKALAVFEGGVHLIETFTVPDADTVIKELS--FLKEKGAI 65

Query: 97 GAGTVMTTLEAKAIDAGAKFLVSPHFDSDIANLANENKVYFPGCATATEIVVARKYKC 156
 GAGTV + E ++A+++GA+F+VSPH D +I+ E V+Y PG T TE+V A K
 Sbjct: 66 GAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTLVKAMKLGH 125

Query: 157 QIIKLFPGGVVGPFGFIKDIHGPIPDVDLMPSGGVSVSNVVEWRKAGAVAVGVGSALSSKV 216
 I+KLFEF VVGP F+K + GP P+V +P+GGV++ NV EW KAG +AVGVGSAL
 Sbjct: 126 TILKLFPGGVVGPQFVKAMKGPFPNVKFPVPIGGVNLNDNVCEWFKAGVLAVGVGSALVKGT 185

Query: 217 ATEGYDSVTKIAKQFVSAL 235
 D V + AK FV +
 Sbjct: 186 P-----DEVREKAKAFVEKI 200

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1251> which encodes the amino acid sequence <SEQ ID 1252>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1039(Affirmative) < succ>

-488-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5 Identities = 82/204 (40%), Positives = 132/204 (64%)

Query: 32 MLNQLKNNYFFAVIRGKSSSEDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNN 91
 +L +LK N V+RG+SSE+AL + +I GGI+ IEVT++ P AS+VI QL++ FK +
Sbjct: 6 ILTCLKANRLVLVVRGESSEBALACSLASIEGGIKTIEVTYTNPFASEVIGQLAERFKED 65

10 Query: 92 KEIIVGAGTVMTTLEAKEAIDAGAKFLVSPHFDSDIANLANENKVVYFPGCATATEIVVA 151
 E+++GAGTV+ A++AI AGA+F+V P+F+ +A + + + Y PGC T E+V A
Sbjct: 66 PEVLIGAGTVLDDVTARQAILAGAQFIVGPNFNRAVALICHRYSHIPYLPGCMTVNEVVTA 125

15 Query: 152 RKYKQCIKLFPGGVVGPFIKDIHGPIPDVDLMPSGGVSVSNNVEWRKAGAVAVGVGSA 211
 + ++K+FPG VG FI+ I P+P V++M +GGVS N+ +W AG +G+G
Sbjct: 126 LESGVDMVKIFPGSTVGISFIRAIKSPLPQVEVMVTGGVSSDNLKDWLAAGVDVLGIGGE 185

20 Query: 212 LSSKVATEGYDSVTKIAKQFVSAL 235
 + + + Y+ +TK A ++ +L
Sbjct: 186 FNQLASQKQYNLITKKAHYIKSL 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 385

A DNA sequence (GBSx0416) was identified in *S. agalactiae* <SEQ ID 1253> which encodes the amino acid sequence <SEQ ID 1254>. This protein is predicted to be pyruvate dehydrogenase complex repressor. Analysis of this protein sequence reveals the following:

30 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2827 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12044 GB:Z99105 similar to transcriptional regulator (GntR family) [Bacillus subtilis]

40 Identities = 67/225 (29%), Positives = 119/225 (52%), Gaps = 17/225 (7%)

Query: 3 RPLVEQTADRLHLILEREYPVGA KLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQ 62
 + L +Q +R++HL+ + G KLP E EL + L V R +REA+ SL T ++ +
Sbjct: 16 KTLAKQVIERIVHLLSSGQLRAGDKLPTEMELMDILHVS RPFVLRREALSSLET LGVITRKT 75

45 Query: 63 GSGTYISSKKGVS EDP L GFS LIKDTDR L TSD L F E L R L L L E P R I A E L V A Y R I T D D Q L Q L L E 122
 GTY + K G+ P L TD L + + E R+ LE + + A +I +++LQ L+
Sbjct: 76 RGTTYFNDKIGM--QPFSVMLALATDNLP A-I I E A R M A L E L G L V T I A A E K I N E E E L Q R L Q 132

50 Query: 123 KLVGDIEDAV--HAGDPKHL L L D V E F H S M L A K Y S G N I A M D S L L P V I N Q S I H L I N A N Y T N R 180
 K + DI ++ H G+ D E F H ++A + N ++ ++ QS+ + +A ++
Sbjct: 133 KTIDDIANSTDNHYGE-----ADKEFHRIIALSANNP V V E G M I ----Q S L L I T H A K I D S Q 183

55 Query: 181 ---QMKSDSLEAHREIIKAIREKNPVA A H D A M L M H I M S V R R S A L K 222
 + + ++E H++I A+ +++P A H M H+ VR LK
Sbjct: 184 IPYRERDVTVEYHKIYDALAQ R D P Y K A H Y H M Y E H L K F V R D K I L K 228

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1255> which encodes the amino acid sequence <SEQ ID 1256>. Analysis of this protein sequence reveals the following:

-489-

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2161(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 24/51 (47%), Positives = 35/51 (68%)

Query: 22 YPVGAKLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQSGTYISSKK 72
 +P+G++LP+E LAE V R T+R+A+ L ILE R GSGTY++S +
 Sbjct: 30 WPIGSRPLPSERHLAEHFTVSRMTLRQAITLLVEEGILERRIGSGTYVASHR 80

15

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 386

20 A DNA sequence (GBSx0417) was identified in *S.agalactiae* <SEQ ID 1257> which encodes the amino acid sequence <SEQ ID 1258>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2178(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9701> which encodes amino acid sequence <SEQ ID 9702> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA58911 GB:X84105 gluceronidase [synthetic construct]

Identities = 258/602 (42%), Positives = 357/602 (58%), Gaps = 31/602 (5%)

35 Query: 23 MLYPLLTKTTRNTYDLGGIWNFKLGEHNPN-----ELLPSDEVMVIPTSFNDLMVSKEK 75
 ML P+ T TR L G+W F L N L + +P SFND +
 Sbjct: 1 MLRPVETPTREIKKLDGLWAFSLDRENCIGDQRWWESALQESRAIAVPGSFNDQFADADI 60

40 Query: 76 RDYIGDFWYEKVEVPKVSDEEMVLRFGSVTHQAKIYVDGVLVGEHKGGFPPFEVLVPE 135
 R+Y G+ WY++ + +PK + +VLR+ +VTH K++V+ V EH+GG+TPFE V
 Sbjct: 61 RNYAGNVWYQREVFIPKGWAGQRIVLRFDVAVTHYGVWVNNQEVMEHQGGYTPFEADVTP 120

45 Query: 136 CKYNNEKIKVSI CANNVLDYITLPVGNYSIIQEDGSIKKKVRNFDFFNAGVHRPLKL 195
 + +++++C NN L++ T+P G I E+G KKK DFFNYAG+HR + L
 Sbjct: 121 YVIAGKSVRITVCVNNELNWQTIPPGMV--ITDENG--KKKQSYFHDFFNAGIHRSVML 176

50 Query: 196 MIRPKNHIFDITITSRLSDDLQSAHLFLVETNQKVDEVRI SVFDEDNKLIV--GETKDSR 253
 P + DIT+ + ++ D A + + V N +V + + D D ++V G+
 Sbjct: 177 YTTPNTWDDITVTVTHVAQDCNHASVDWQVVAN---GDVSVELRDADQQVVATGQGTSGT 233

Query: 254 LFLSDVHLWEVLNAYLYTARVEIFVDNQLQDVYEENFGLREIEVTNGQFLLNKPIYFKG 313
 L + + HLW+ YLY V + D+Y G+R + V QFL+N KP YF G
 Sbjct: 234 LQVNVNHLWQPGEGYLYELCVTAKSQTEC-DIYPLRVGIRSVAVKGEQFLINHKPFYFTG 292

55 Query: 314 PGKHEDTFINGRGLNEAANLMDLNLKDMGANSFRTSHYPYSEEMRLADRMGVLVIDEV 373
 FG+HED + G+G + + D L+ +GANS+RTSHYPY+EEM+ AD G++VIDE
 Sbjct: 293 FGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVVIDET 352

Query: 374 PAVGLFQNFNASLDLS-----PKDNGTWNLM--QTKAAHEQAIQELVKRDKNHPSVVMW 425

-490-

AVG FN SL + PK+ + + +T+ AH QAI+EL+ RDKNHPSVVMW
 Sbjct: 353 AAVG----FNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMW 408

5 Query: 426 VVANEPASHEAGAHDYFEPLVKLYKDLDPQKRFTLVNLMATPDRDQVMDLVDVCLNR 485
 +ANEP + GA +YF PL + + LDP RP+T VN++ D + DL DV+CLNR
 Sbjct: 409 SIANEPDTRPQGAREYFAPLAEATRKLDP-TPITCVNMFCDAHTDTISDLFDVCLNR 467

10 Query: 486 YYGWYVDHGDLTNAEVGIRKELLEWQDKFPDKPIITEYGADTLPLHSTWNIPTTEEFQ 545
 YYGWYV GDL AE + KELL WQ+K +PIITEYG DTL GLHS + ++EE+Q
 Sbjct: 468 YYGWYVQSGDLETAEKVLEKELLAWQEKI-HQPIITEYGVDTLAGLHSMYTDWSEEYQ 526

Query: 546 CDFYEMSHRVFDGIPNLVGEQVWNFADFETNLMILRVQGNHKGFLSRNRQPKQVVKEFKK 605
 C + +M HRVFD + +VGEQVWNFADF T+ ILRV GN KG+F+R+R+PK +K
 15 Sbjct: 527 CAWLDMYHRVFDVSAVVGQVWNFADFATSQGILRVGGNKGIFTDRPKSAFFLLQK 586

Query: 606 RW 607
 RW
 Sbjct: 587 RW 588

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1259> which encodes the amino acid sequence <SEQ ID 1260>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -4.04 Transmembrane 1131 -1147 (1130 -1147)

----- Final Results -----
 bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF97242 GB:AF282987 beta-galactosidase precursor [Streptococcus pneumoniae]
 Identities = 303/921 (32%), Positives = 463/921 (49%), Gaps = 86/921 (9%)

35 Query: 5 QKSSEIVT----RTITKPSRATSNVKQEIDMTFDSKEQTVTGYYHYIDQ--BGRKQPFN 58
 +K E VT + KP ++ + ++ ++Q E RK FN
 Sbjct: 96 KKEDEAVTPKEEKVSAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDNF 155

40 Query: 59 QGWRFLMADVACAQDPSFDDSNWQVIHLPHDFSLTQPYTRNGEA--ESAYKLGGVGWYR 115
 Q W F L A+ A P D S W+ + LP+D+S+ + A E GG WYR
 Sbjct: 156 QNWYFKLNANSKEAIKPDADVSTWKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAWYR 215

45 Query: 116 HYLVLDEVLAGCHVAITFEGSYMETEIVNGQFIGKHLNGYQEFYDISDVVTF-GAENL 174
 LDE +V +TF+G YM++++YVNGQ +G + NGY +F+YDI+ + G EN+
 Sbjct: 216 KTFKLEKDLKKNVRLTFDGVYMDSQVYVNGQLVGHYPNGYNQFSYDITKYLQKDGREN 275

Query: 175 LAVRVENKVPSSRWYSGSGLYREVSLSVLPQLHFVADQVAMTLADTAVQEKGGQKVDLRF 234
 +AV NK PSSRWYSGSG+YR+V+L V ++H + + Q+ G+ + +
 50 Sbjct: 276 IAVHAVNKQPSRWYSGSGIYRDVTLQVTDKVHVEKNGTTLTPKLEEQQHGKVVETHVTS 335

Query: 235 ALNQSIQTCHYQLSLCLWEQSHCSKDKLLYQETEVLADLAFQRQYGLT--LSLEELQL 292
 + + H ++ E + + L L L +E +L
 Sbjct: 336 KIVNTDDKDHELVA---EYQIVERGGHAVTGLVRTASRTLKAHESTSLDAILEVERPKL 391

55 Query: 293 WSP--DNPHLYDLELTLYYQQGVIDCFLETGFRQLTFMANQGLFVNGRAVKLGVCVCLHH 350
 W+ D P LY+L +Y GQ++D G+R + N+G +NG +K GV LHH
 Sbjct: 392 WTVLNDKFPALYELITRVYRDGQLVDAKKDLFGYRYHWTNNEGFSNLNGERIKFHHGVSLLHH 451

60 Query: 351 DQGGGLGACAYEDALARQLVLLKDMGANTIRSTHNPSSPKLRQLANRLGFFVIEEAFDTWT 410
 D G LGA A R+L +K+MG N+IR+THNP+S + Q+A LG V EEAFTW
 Sbjct: 452 DHGALGAENYKAEYRRLKQMKEMGVNSIRTTNPASEQTLQIAAELGLLVQEEAFTWY 511

Query: 411 YAKNGNVNDFSNYFHQTIGTENANYLQVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIG 470
 65 K D+ +F + A ++ W+ + + MV KN+P++ MWSIG

-491-

Sbjct: 512 GGG--KPYDYGRFFEKDATHEPARKGEK-----WSDFDLRTMVERGKNNPAIFMWSIG 562

Query: 471 NELMEGFSADVSHYPELTRQMCQWITAITDSRPITFGDNKLKEADFC--WHEEVSQMATLL 529
NE+ G + +H +++ + I +D +R +T G +K + + HE+++

5 Sbjct: 563 NEI--GEANGDAHSLATVKRLVKIKVDKTRYVTMGADKFRFGNGSGGHEKIA----- 614

Query: 530 SQLDHPQGLIGLNYADGKDYDRLHEEHSDWLLYGSSETVSATISR--AYYKETKKVLDS--- 585
+LD +G NY++ +Y L +H WL+YGSET SA +R +YY+ +++ S

10 Sbjct: 615 DELD----AVGFNYSE--DNYKALRAKHPKWLIIYGETSSATRTRGSYRPERELKHSNGP 669

Query: 586 --GYHLTSYDHAKVDWGAFASQAWYDTITRDFV--AGECVWTGFDYLGEPWPWNKIDSGV 641
Y + Y + +V WG A+ +W T RD AG+ +WTG DY+GEPTPW+ +

15 Sbjct: 670 ERNYEQSDYGNDRVWGKTATASW--TFDRDNAGYAGQFIWTGTDIYIGEPTPWHNQQTTP 727

Query: 642 VGLWSPKPNAYFGILDTAGFPKDSYFYQSQW--AQCQTTLHLLPVWQKD-----QLCFD 694
V K++YFGI+DTAG EK +Y YQSOW + + +HLLP W + D

Sbjct: 728 V-----KSSYFGIVDTAGIPKHDIFYLYQSQWVSVKKKPMVHLLPHWNWENKELASKVAD 781

20 Query: 695 EQGLEVEVVYSNAASVQIMFEDEQNLTDYGRKAFHTYSTPTGHTYQLYQGADAANKPHE 754
+G + V YSNA+SV+L N G K F+ T G TYQ +GA+A

Sbjct: 782 SEGKIPVRAYSNAASSVELFL-----NGKSLGLKTFNKKQTS DGRTYQ--EGANA-----N 829

Query: 755 NLYLTWRVPYQKGLLRAVAYDISGKSIPKTSGRSQVRTYGSVAKLSWKAFAEPIIDAPW-E 813
LYL W+V YQ G L A+A D SGK I R ++ T G A + + I A +

25 Sbjct: 830 ELYLEWKVAYQPGTLEAIARDESCKEI---ARDKITTACKPAAVRLIKEDHALAADGKD 885

Query: 814 LLYLDLSLDSRGELVSHAQDWLQVQVEGPARLLALDNGNPTDHTFYQEP-----LRQAY 868
L Y+ ++DS+G +V A + ++ Q+ G +L+ +DNG Y+ +R+A+

30 Sbjct: 886 LTYIYYEIVDSQGNVVPITANNLVRFLQHGQQLVGVNDGEQASRERYKAQADGSWIRKAF 945

Query: 869 GKKLLAILALTGEAGHIKVT 889
GK +AI+ T +AG +TA

35 Sbjct: 946 NGKGVAIVKSTEQAGKFTLTA 966

An alignment of the GAS and GBS proteins is shown below:

Identities = 98/414 (23%), Positives = 175/414 (41%), Gaps = 64/414 (15%)

Query: 54 LPSDEVMIPTSFNDLMVSKERDYGDFWYKIVKPKVSEDEEMVLRFGSVTHQAKIY 113
LP D + P + N S K +G WY + + +V + + F + +IY

40 Sbjct: 86 LPHDFSLTQPYTRNGEAEASAYKLGVG--WYRHYLVLDVLGCHVAITFEQSYMETEY 143

Query: 114 VDGVLVGEHKGFTPFVVLVPECKYNNEKIKVSIANNVLDYTTLPVGNYSIIQEDGSI 173
V+G +G+H G+ F + + V+ A N+L +

45 Sbjct: 144 VNGQFIGKHLNGYQFTYDISDV-----VTFGAENLLAVR-----V 179

Query: 174 KKKVRENFDFFNYAGVHRPLKLMIRPKNHIFDITITSRLSDDL-----QSADLHFLVET 227
+ KV + +++ +G++R + L + P+ H + L+D Q DL F +

50 Sbjct: 180 ENKVPSS--RWYSGSGLYREVLSVLPLQHFVADQVAMTLADTAVQEKQKQVDLRFALNQ 238

Query: 228 NQKQDEVRIISVF-----DEDNKLVGETKDS-----RLFLSDVHLWEVLNA 267
+ + ++S+ +D KL+ + + L L ++ LW N

Sbjct: 239 SIQTCHYQLSLCLWEQSHCSKDKKLLYQETEVPLADLAFQRYGLTSLSEELQLWSPDNP 298

55 Query: 268 YLYTARVEIFVDNQLQDVVEENFGLREIE-VTNGQFLNRPKIYFKGFGKHEDTFINGRG 326
+LY + ++ Q+ D + G R++ + N +N + + KG H D G

Sbjct: 299 HLYDLELTLYYQGGVIDCFLETGFRQLTFMANQGLFVNGRAVKLKGVCLHHDQGGGLGAC 358

Query: 327 LNEAANLMDLNLKDMGANSFRTSHYPYSEEMRLADRMGVLVIDEVPAVGLFQ---NFN 383
E A L LKDMGAN+ R++H P S ++ +LA+R+G VI+E + N N

60 Sbjct: 359 AYEDALARQLVLLKDMGANTIRSTHNPSPKLRQLANRLGFFVIEEAFDTWTYAKNGNVN 418

Query: 384 ASLDLSPKDNCTWN---LMQTKAAH---EQAIQELVKRDKNHPSSVVMWVVANE 430
+ + GT N L + ++ + +I+ +V KN PSV+MW + NE

65 Sbjct: 419 DFSNYFHQTIGTENANYLQVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIGNE 472

-492-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 387

A DNA sequence (GBSx0418) was identified in *S. agalactiae* <SEQ ID 1261> which encodes the amino acid sequence <SEQ ID 1262>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.53    Transmembrane 197 - 213 ( 197 - 213)

----- Final Results -----
          bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9699> which encodes amino acid sequence <SEQ ID 9700> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
Identities = 115/342 (33%), Positives = 180/342 (52%), Gaps = 16/342 (4%)

Query: 14  KIISLGEVLLRLSPQYHTLMQANHLKCQFGGSELNVLASLAQLGYHVGVLVSALPDNDLG 73
          K+++ GE++LRLSP +  + Q +  +GG+E NV A LAQ+G  V+ LP+N LG
Sbjct: 2   KVVTFGEIMLRLSPPDHKRIFQTDSEFVTVGGAEANVAAFLAQMGDAYFVTKLPNNPLG 61

Query: 74  KMASQFILSQQISPAAIKKEGRLGIYYEQQGFSVRTNKVIYDRNYSSFWESTLSDYDFT 133
          A+ + + I + R+GIY+ E G S R +KV+YDR +S+ E+ D+D+
Sbjct: 62  DAAAGHLRKFGVKTDYIARGGNRIGIYFLEIGASQRPSKVVDRAHSAISEAKREDFDWE 121

Query: 134 SIFKGVDFWVHSGITPALT KDLYEVTRFLMTKAKEGGVKVSDLNFRSLWSSFQEAREQ 193
          I G WFH SGITP L K+L + + A E GV VS DLN+R LW+ +EA++
Sbjct: 122 KILDGARWVHFGSGITPPLGKELPLILEDALKVANEKGVTVSCDLNYRRLWTK-BEAQKV 180

Query: 194 LSPLLGLLDVCFGLEPIYLAGESEDLKDELGLSRPYLDI-----ELLEKITQKIVQEV 246
          + P + +DV L ED++ LG+S LD+ E KI +++ ++Y
Sbjct: 181 MIPFMEYVDV-----LIANEEDIEKVLGISVEGLDLKTGKLNREAYAKIAEEVTRKY 232

Query: 247 GLDYIAFTQREMEYTNQYMLKSYLYHNNMLYQTDKTVGEVLDVRVGTGDAFAAGLIHALLE 306
          + T RE ++ N + +++ + ++DRVG GD+FA LI+ L
Sbjct: 233 NFKTVGITLRESISATVNYWSVMVFENGQPHFSNRYEIHIVDRVAGDSFAGALIYGLSM 292

Query: 307 KETPQRALEIAMATFKYKHTIQGDINIMTRDDIAYLIEKETN 348
          Q+ E A A KHTI GD +++ ++I L T+
Sbjct: 293 GFDSQKAEFAAAASCLKHTIPGDFVVLSEIEIEKLAGSATS 334
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1263> which encodes the amino acid sequence <SEQ ID 1264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0708(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/319 (34%), Positives = 168/319 (51%), Gaps = 7/319 (2%)
```

-493-

Query: 12 MAKIISLGEVLLRLSPPOYHTLMQANHLKCQFGGSEINVLASLAQLGYHVGLVSALPDND 71
 M+K++ +GE L+R+SP Q+ L A + FGGSE+N+ +L G L +ALPDN
 Sbjct: 14 MSKLLLVGEPLIRVSPNQFQPLTNACEAQLFFGGSEVNIARTLGGFGLEARLFTALPDNP 73

5 Query: 72 LGKMASQFILSQQISPAAIKKEGRLGIYYEQQGFSVRTNKVIYDRNYSSFWESTLSDVD 131
 +G QF+ + + + R+G+YY E GF R ++V YDR SSF D
 Sbjct: 74 VGHAHFHQFLKQSGVMSLTAWQGHVGLYYLENGFGCRASQVYVDRCGSSFSALDKDSDL 133

10 Query: 132 FTSIFKGVDFWVHSGITPALT KDLYEVTRFLMTKAKEGGVKV SIDLNPRESLWSSSQEAR 191
 +IF+G+ FH SGI+ AL K ++ L+ +AK+ + +S DINFR S+ + +A+
 Sbjct: 134 LAAIFEGISHFHFSGISLALGKKTQDLIEVLVREAKKRDICISFDLNRSSM-IAVADAK 192

15 Query: 192 EQLSPLLGLLDVCFGLEPTIYLAGESEDLKDELGLSRFYLDIELLEKITQKIVQEYGLDYI 251
 S D+ FG+EP+ L + D+ D R D + + + Q Y L I
 Sbjct: 193 RLFSHFAQYADIIFGMEPLLLDSDDFDMFD----RKKADTTTIRERLAGLYQRYQLQAI 247

20 Query: 252 AFTQREMEYTNQYMLKSYLYHNNMLYQTDKGTGVEVLDRVGTGDAFAAGLIHALLEKETPQ 311
 T+R + K+Y Y + Y++ + VL RVG+GDAF AGL++ LLE Q
 Sbjct: 248 YHTERSNDAGGSNHFKAYAY-DRQFYESCEVTTFVLQRVGSGDAFVAGLLYQLLEGNEKQ 306

Query: 312 RALEIAMATFKYKHTIQGD 330
 R L+ A+AT K T+ D
 Sbjct: 307 RNLDFAVATASLKCTVAED 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 388

A DNA sequence (GBSx0419) was identified in *S.agalactiae* <SEQ ID 1265> which encodes the amino acid sequence <SEQ ID 1266>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.17 Transmembrane 5 - 21 (5 - 21)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 389

45 A DNA sequence (GBSx0420) was identified in *S.agalactiae* <SEQ ID 1267> which encodes the amino acid sequence <SEQ ID 1268>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -12.05 Transmembrane 198 - 214 (191 - 220)
 INTEGRAL Likelihood = -11.68 Transmembrane 446 - 462 (437 - 467)
 INTEGRAL Likelihood = -9.55 Transmembrane 94 - 110 (91 - 116)
 INTEGRAL Likelihood = -7.43 Transmembrane 291 - 307 (283 - 309)
 INTEGRAL Likelihood = -4.88 Transmembrane 265 - 281 (257 - 282)
 INTEGRAL Likelihood = -4.62 Transmembrane 321 - 337 (318 - 339)
 55 INTEGRAL Likelihood = -3.93 Transmembrane 406 - 422 (405 - 426)
 INTEGRAL Likelihood = -1.59 Transmembrane 121 - 137 (121 - 137)

-494-

INTEGRAL Likelihood = -1.12 Transmembrane 345 - 361 (345 - 362)
 INTEGRAL Likelihood = -0.48 Transmembrane 43 - 59 (43 - 59)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB13641 GB:Z99113 similar to H+-symporter [Bacillus subtilis]
 Identities = 105/452 (23%), Positives = 182/452 (40%), Gaps = 37/452 (8%)

Query: 36 IYLFTFMFVTVFSTGVLGSAAIFVSQIMGYIRIFDGFIDPAIGIMIDKTDTKFGKYPIL 95
 IY ++ +F T V G +A + +RI D DP IG ++D+T+++F ++RP L

15 Sbjct: 27 IYATVSTYLLFFYTDVFGLSAAAAGTMFLVVRIDALADPFITIVDRITNSRFARFRPYL 86

Query: 96 IIGNVITALSLIFLLALRGVDENIRFPLFILVLIHKIGYSMQQTITKAGQTALTNDPKQ 155
 + G A + L L + ++ I +G S+ T ALT+

20 Sbjct: 87 LFG----AFPFVILAILCFTTPDFSDMGKLIYAYITYVGLSLTYTTINVPYGALTS-AMT 141

Query: 156 RPIFINIVDAVMTTSLMTGGQFVSVFLVPKFGNFTPQFFNVLI FG TILISAILAIV--AI 213
 R +V L +V F VP + G L IL ++ +

25 Sbjct: 142 RNNQEVVSITSVRMLFANLGLLVAFVPLLAAYLSDTSGNESLQWQTMGILGMIGGCL 201

Query: 214 IGIWAKDRKEFFCLGENTQKTALKDYWKVLKGNKPLQILSIAAALVKFAIQFFGDSV-VM 272
 + K KE L ++ +K D ++ + N+PL +LSI ++ F + +SV +

30 Sbjct: 202 LIFCFKSTKERVTLQKSEKIKFTDIFEQFRVNRPLVLSIFFIII-FGVNSISNSVGIY 260

Query: 273 VLLFGI----LFGNYALSGQFSLLEFIVPGVIINILFSTIARKKGLRFSYVRAIQIGMIGL 328
 + + + L Y L G L I+P I L + +KK L + A+ + +IGL

35 Sbjct: 261 YVTYNLEREDLVKWYGLIGSLPALVILP--FIPRLHQFLGKKLLNY----ALLNIIGL 314

Query: 329 LAFGAVLYVGKFGDLSTSLNLYTILFTVTNIIARYASQAPASLVLTMGADISDYETSES 388
 LA L + N+Y IL V +IA S + + + +Y +

35 Sbjct: 315 LAL-----LFVPPSNVYLIL--VCRLIAAAGSLTAGGYMWALIPETIEYGEYRT 361

Query: 389 GRYVSGMIGTIFSLTDSIASSFAPMVVGFVLGIGFSKSFPTTETPLPPDLKMAAISILV 448
 G+ + G+I I + +V G VL G+ P M +

40 Sbjct: 362 GKRMGGLIYAIIGFFPKFGMALGGVVPGLVLDKFGY-----VANQAQTPAALMGILITTT 416

Query: 449 AIPFIALSIALLLMKFYKLDKEEMVRIQEKIQ 480
 IP L +AL+ + FY LD+++ + +++

45 Sbjct: 417 IIPVLLVLALIDINFYNLDEKKYKYNMVRELE 448

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 390

50 A DNA sequence (GBSx0422) was identified in *S.agalactiae* <SEQ ID 1269> which encodes the amino acid sequence <SEQ ID 1270>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-495-

>GP:AAB17663 GB:U31175 D-specific D-2-hydroxyacid dehydrogenase [*S. aureus*]
Identities = 165/331 (49%), Positives = 231/331 (68%), Gaps = 1/331 (0%)

5 Query: 1 MMKLKVFNVREEEATLAQDWANRNHVELSMSEGPLTLETVNEVEGFDGIANAQIEPLDDA 60
M K+ F R+ E +A +W +N+VE++ S+ L+ TV++++ +DG+ Q L++
Sbjct: 1 MTKIMFFGTRDYEKEMALNWGKKNVETTSKELLSSATVDQLKDYDGVITMQFGKLEND 60

10 Query: 61 IYPLLKEMGIKQIAQRSAGVDMYNLELAKQHGIISNVPSYSPEIAEFTVTIALNLIRK 120
+YP L+ GIKQIAQR+AG DMY+L+LAK+H I+ISNVPSYSPE+IAE++V+IAL L+R+
Sbjct: 61 VYPKLESYGKQIAQRTAGFDMDLAKKHIVISNVPSYSPETIAEYSVSIALQLVRR 120

15 Query: 121 VELIRANVREQNFSWTLPIRGRVLGNMTVAIIGTGRIGLATAKIFKGFGRVIGYDIYHN 180
I V+ +F+W I + + NMTVAIIGTGRIG ATAKI+ GFG + YD Y N
Sbjct: 121 FPDIERRVQAHDFTWQAEIMSKPVKNMTVAIIGTGRIGAATAKIYAGFGATTAYDAYPN 180

20 Query: 181 PMADGILEYVNSVEEAEVEADLVSLHMPPTAENTHLENLDMFKQFKGAILMNMARGALV 240
D L Y +SV+EA+++AD++SLH+P E+ HLF+ MF KKGAIL+N ARG++
Sbjct: 181 KDLD-FLTYKDSVKEAIKDADIISLHVPANKESYHLFDKAMFDHVKKGAILVNAARGAVI 239

25 Query: 241 ETKDLLLEALDQGLLEGAGIDTYEFEGPYIPKNCQGQDISDKDFLRLLINHPKVIYTPHAAY 300
T DL+ A++ G L GA IDTYE E Y + +DI DK L LI H +++ TPH A+
Sbjct: 240 NTPDLIAAVNDGTLGAAIDTYENEAAYFTNDWINKDIDDKTLELIEHERILVTPHIAF 299

Query: 301 YTDEAVKNLVEGALNACVEVIETGTTTKVN 331
++DEAV+NLVEG LNA + VI TGT T++N
Sbjct: 300 FSDEAVQNLVEGGLNAALSVINTGTCTETRLN 330

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
30 vaccines or diagnostics.

Example 391

A DNA sequence (GBSx0423) was identified in *S.agalactiae* <SEQ ID 1271> which encodes the amino acid sequence <SEQ ID 1272>. Analysis of this protein sequence reveals the following:

35 Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
45 vaccines or diagnostics.

Example 392

A DNA sequence (GBSx0424) was identified in *S.agalactiae* <SEQ ID 1273> which encodes the amino acid sequence <SEQ ID 1274>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

50 Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood =-12.90 Transmembrane 64 - 80 (53 - 89)

----- Final Results -----

-496-

bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9325> which encodes amino acid sequence <SEQ ID 9326> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
 pallidum]
 10 Identities = 33/91 (36%), Positives = 55/91 (60%), Gaps = 1/91 (1%)
 Query: 1 MANVLAKPKIMLFMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGWSV 60
 M N + P + +P++ + + G+L LFN+QGTPASAGFG GL+GPINA L V
 15 Sbjct: 250 MPNWIRYPILNIPLLNGLVCGVLAWLFLNLQGTPASAGFGFGLVGPINAYRLMAYTPMV 309
 Query: 61 MNMLLIIIIIFVAAPIILNFIFNYLFIKVLKI 91
 +L ++ FV + + ++ +++ + LK+
 20 Sbjct: 310 RAGILFLVYFVLS-FLAAYLIDFILVDRLKL 339

- 20 A related DNA sequence was identified in *Spyogenes* <SEQ ID 1275> which encodes the amino acid sequence <SEQ ID 1276>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have a cleavable N-term signal seq.
 25 INTEGRAL Likelihood = -12.31 Transmembrane 141 - 157 (133 - 166)
 INTEGRAL Likelihood = -6.00 Transmembrane 92 - 108 (88 - 112)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
 pallidum]
 35 Identities = 63/178 (35%), Positives = 107/178 (59%), Gaps = 10/178 (5%)
 Query: 2 IGQGIASLLGLQPILMSLLIAMIFCFLIVSPITTVGIALAINLSGIGSGAASFG----- 55
 +G+ IA+ + LQP+LMS+L++M F +I+SP+++V + +A+ L+G+ SGAA+ G
 40 Sbjct: 164 VGRVIATFTIALQPLLSILLSMSFSLIIISPVSVAVGIAVGLTGLASGAANIGVSSCAM 223
 Query: 56 -LCLAGWAVNSKGTSLAHVLRSPKISMANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTP 114
 L + VN G LA + K+ M N + P + +P+L + V GV+ +FN+QGTP
 Sbjct: 224 TLIVGTMRVKNIGVPLAMFAGAMKMLMPNWIRYPILNIPLLNGLVCGVLAWLFLNLQGTP 283
 45 Query: 115 ASAGFGISGLIGPINALNLAKGGWCP-VNILLIIIIIFVGAPIVLNMIFNYLFIKVLKV 171
 ASAGFG GL+GPINA L + P V ++ +++ + + +++ + LK+
 Sbjct: 284 ASAGFGFGLVGPINAYRLM--AYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKL 339

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 86/101 (85%), Positives = 96/101 (94%)
 Query: 1 MANVLAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGWSV 60
 MANVL+KPKIMLPM+ SAA+LG++GA+FNIQGTPASAGFGISGLIGPINALNLAKGGW
 55 Sbjct: 81 MANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTPASAGFGISGLIGPINALNLAKGGWCP 140
 Query: 61 MNMLLIIIIIFVAAPIILNFIFNYLFIKVLKIIDPMDYKLDI 101
 +N+LLIIIIIFV API+LN IFNYLFIKVLK+IDPMDYKLDI
 Sbjct: 141 VNILLIIIIIFVGAPIVLNMIFNYLFIKVLKVIDPMDYKLDI 181

-497-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 393

A DNA sequence (GBSx0426) was identified in *S.agalactiae* <SEQ ID 1277> which encodes the amino acid sequence <SEQ ID 1278>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.58	Transmembrane	148 - 164 (145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 (25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 (62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 (122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 (96 - 112)

----- Final Results -----

bacterial membrane	---	Certainty=0.3633(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9735> which encodes amino acid sequence <SEQ ID 9736> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9697> which encodes amino acid sequence <SEQ ID 9698> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema pallidum]
Identities = 61/158 (38%), Positives = 92/158 (57%)

Query: 24 KSFIMNVNLGLGTVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMGLVIGIM 83
+ F+M +LNG + G VI L+P AI GEL +AL P+ FA L + +IG +
Sbjct: 9 RQFMMKILNGSSAGIVIGLVPPAIAAGELFRALAPLSPLFAALYHVLPPIQFSVPALIGTL 68

Query: 84 VGLNFKFNPISASLGLAVMPAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFLSD 143
VGL F + + A+L + A G T GA ++ G GD+IN+ + +AL ++L++ L
Sbjct: 69 VGLQFHCSAPEVATLAFVSVIASGNVTLQNGAWLITGIGDVINVMLISALAILVRALRG 128

Query: 144 KTKSFILIVIPTVTLVGGVGHVLLPYVKMITTMIGQ 181
K S T+I +P + ++ GGVG LPYVKMIT +G+
Sbjct: 129 KLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFGVR 166

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1279> which encodes the amino acid sequence <SEQ ID 1280>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.06	Transmembrane	314 - 330 (301 - 335)
INTEGRAL	Likelihood = -11.30	Transmembrane	185 - 201 (178 - 215)
INTEGRAL	Likelihood = -8.01	Transmembrane	22 - 38 (11 - 42)
INTEGRAL	Likelihood = -3.29	Transmembrane	266 - 282 (265 - 285)
INTEGRAL	Likelihood = -2.66	Transmembrane	141 - 157 (141 - 159)
INTEGRAL	Likelihood = -2.13	Transmembrane	53 - 69 (53 - 69)
INTEGRAL	Likelihood = -1.33	Transmembrane	114 - 130 (113 - 131)
INTEGRAL	Likelihood = -0.80	Transmembrane	206 - 222 (206 - 222)

----- Final Results -----

bacterial membrane	---	Certainty=0.6222(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

-498-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
pallidum]
Identities = 137/346 (39%), Positives = 217/346 (62%), Gaps = 14/346 (4%)

Query: 12 FMNKVLGTAIAIVVALIPNAILATFLKPLLP-NMAAEFLHIVQVFQFFTPIMAGFLIG 70
FM K+L G++ IV+ L+P AI + L P + A H+V QF P + G L+G

10 Sbjct: 11 FMMKILNGSSAGIVIGLVPPAIAAGELFRALAPLSPLFAALYHVLPPIQFSVPALIGTLVG 70

Query: 71 QQFKFNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMITASLAVLA 130
QF + + + + I SG + G + + GIGD+IN+M+ ++LA++

15 Sbjct: 71 LQFHCSAPEVATLAFVSVIASG-----NVTLQNGAWLITGIGDVINVMLISALAIL 122

Query: 131 VKYFGNKFGLSLTIILLPITIGTGVGYIGWKFLPYVSVYVTTLIGQGINSFTTLQPILMSIL 190
V+ K GSLTII LP+ + G +G LPYV +T +G+ I +F LQP+LMSIL

Sbjct: 123 VRALRGKLGSLTIIALFPVIVAVVAGGVGSFSLPYVKMITLFEVGRVIATFIALQPLMSIL 182

20 Query: 191 IAVAFSLIIVSPISTVAIGLAIGLNGMAAGAASMGIASTA AVLWATLKVNKSGVPPIAIA 250
+++FSLII+SP+S+VA+G+A+GL G+A+GAA++G++S A L+ T++VNK GVP+A+

Sbjct: 183 LMSFSLIIISPVSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTMRVNKIGVPLAMF 242

25 Query: 251 LGAMKMMMPNFKHPIMAIPMVFTAIISSLTIVPLFNLVGTTPASSGFGVLGVAGVPIAS--L 308
GAMKM+MPN++++PI+ IP++ + + LFNL GTPAS+CFG +G VGPI + L

Sbjct: 243 AGAMKMLMPNWIRYPILNIPLLLLNGLVCGVLAWLFNLQGTTPASAGFGFIGLVGPINAYRL 302

Query: 309 AGGSSIL---IILAWIIVPFAVAFAAHKVKSDILKLYKEDIFVFE 351
+ ++ I+ L + ++ F A+ + D LKLY+ ++F+ E

30 Sbjct: 303 MAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIE 348

An alignment of the GAS and GBS proteins is shown below:

Identities = 65/172 (37%), Positives = 95/172 (54%), Gaps = 9/172 (5%)

35 Query: 19 EKQTTKSFIMNVNLGLALGTIVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMGL 78
+K+T SF+ VL G A+ V+ LIP AIL +K LLP + A + V +

Sbjct: 5 DKETFSFNMKVLGTAIAIVVALIPNAILATFLKPLLPNMAA-AEFLHIVQVFQFFTPPI 63

40 Query: 79 VIGIMVGLNFKFNPIQSASLGLAVMFAGGAATFLK-----GAIMLKGTGDIINMGIT 130
+ G ++G FKFNP+Q ++G A GA + + G L+G GD+INM IT

Sbjct: 64 MAGFLIGQQQKFNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMIT 123

Query: 131 AALGVLLIQFLSDKTKSFTLIVIPTVTLILLVGGVGHVLLPYVKMITTMIGQG 182
A+L VL +++ +K S T+I++P VG +G LPYV +TT+IGQG

45 Sbjct: 124 ASLAVLAVKYFGNKFGLSLTIILLPITIGTGVGYIGWKFLPYVSVYVTTLIGQG 175

A related GBS gene <SEQ ID 8567> and protein <SEQ ID 8568> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: -13.49
GvH: Signal Score (-7.5): -5.82
Possible site: 48

>>> Seems to have no N-terminal signal sequence

55 ALOM program count: 5 value: -6.58 threshold: 0.0

INTEGRAL	Likelihood = -6.58	Transmembrane	148 - 164 (145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 (25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 (62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 (122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 (96 - 112)
PERIPHERAL	Likelihood = 1.85	51	

60 modified ALOM score: 1.82

*** Reasoning Step: 3

-500-

Query: 120 EVSLIQDGSRRKQLQELLAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLA 179
 L++DGS K+QE + G +DIG++ P+ + + T + +V+ SH LA
 Sbjct: 121 TFQLVEDGSIKVGEGVGDGSLDIGVVLPANEDIFHSFTIVKET-----LMLVVHPSHRLA 176

5 Query: 180 TLPEIELNDLRDYKVASLNEHYMLGEMLPKCRALGFDPHIVFKHNDWEVLIHSLQDLNA 239
 E +L +L+D E ++L + +C GF PHI+++ + W+ + +
 Sbjct: 177 DEKECQLRELKDEPFIFRFEDFVLHNRIMTECIKAGFRPHIIVETSQWDFISEMVSANLG 236

10 Query: 240 VTILPSEFESISQVQDLWCWVPLKDKNNFYPIGIAYRNDTSFS 281
 + +LP + + +PL D + + I +R D S
 Sbjct: 237 IGLLPERICRGLDPEKVKVIPLVDPVIPWHLAIWRKDRYLS 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1283> which encodes the amino acid sequence <SEQ ID 1284>. Analysis of this protein sequence reveals the following:

15 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1101(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 125/160 (78%), Positives = 144/160 (89%)

25 Query: 135 LAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLATLPEIELNDLRDYKV 194
 L++GKIDIGLLSF S R DITIE LQTSTKGYKVSIV+ K HPLA P+++L DL+ YK+
 Sbjct: 1 LSQGKIDIGLLSFLSIRKIDITIELLQTSTKGYKVSIVLLKQHPLAQHPQLKLDLKGyki 60

30 Query: 195 ASLNEHYMLGEMLPKCRALGFDPHIVFKHNDWEVLIHSLQDLNAVTILPSEFESISQVQ 254
 ASLN+HYMLGEMLPKCRALGF+P IVFKHNDWEVLIHSL DLN +TILPS+FES++QV
 Sbjct: 61 ASLNDHYMLGEMLPKCRALGFEPDIVFKHNDWEVLIHSLHDLNLTILPSDFESLNQVD 120

Query: 255 DLCWVPLKDKNNFYPIGIAYRNDTSFSPMIEEFLSLLKTN 294
 35 +L W+PL+DKNNFYPIGIAYR+D SFSP+IEEFLSLLKTN
 Sbjct: 121 NLVWIPLQDKNNFYPIGIAYRDDASFSPVIEEFLSLLKTN 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 395

A DNA sequence (GBSx0429) was identified in *S.agalactiae* <SEQ ID 1285> which encodes the amino acid sequence <SEQ ID 1286>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1833(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 Signal peptide: 1-21

A related GBS nucleic acid sequence <SEQ ID 8569> which encodes amino acid sequence <SEQ ID 8570> was also identified.

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8570 (GBS271) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 8; MW 31.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 6; MW 56.3kDa) and in Figure 62 (lane 10; MW 56.3kDa).

5 GBS271-GST was purified as shown in Figure 210, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 396

A DNA sequence (GBSx0430) was identified in *S.galactiae* <SEQ ID 1287> which encodes the amino acid sequence <SEQ ID 1288>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -6.74    Transmembrane    9 - 25 ( 5 - 28)
    INTEGRAL    Likelihood = -5.84    Transmembrane    97 - 113 ( 92 - 122)
15    INTEGRAL    Likelihood = -5.47    Transmembrane    37 - 53 ( 35 - 61)
    INTEGRAL    Likelihood = -2.55    Transmembrane    220 - 236 ( 220 - 238)
    INTEGRAL    Likelihood = -1.65    Transmembrane    64 - 80 ( 63 - 81)
    INTEGRAL    Likelihood = -1.28    Transmembrane    193 - 209 ( 192 - 209)
20    INTEGRAL    Likelihood = -0.53    Transmembrane    125 - 141 ( 125 - 141)

----- Final Results -----
        bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC73593 GB:AE000155 putative metal resistance protein
[Escherichia coli K12]
Identities = 128/252 (50%), Positives = 186/252 (73%)
30
Query: 5 NSISLMSLLMASSLVLTITLFFSYWQKLNLEKEVIISAIRAVIQLLAVGFLLDYIFGYQNP 64
      ++I+ SL +A LV++ + S+ +KL LEK+++ S RA+IQL+ VG++L YIF +
Sbjct: 13 HNTNESLALALMLVVVAILISHKEKLALEKDILSVSGRAIIQLIIVGYVLKYIFSVDDA 72

Query: 65 IFTALLMLFMIINASYNAAKRGKGINKGFVISFIAIGSGTIITLSVLIFSGILKFVPNQ 124
      T L++LF+ NA++NA KR K I K F+ SFIAI G ITL+VLI SG ++F+P Q+
35 Sbjct: 73 SLTLLMVLFCFNAAWNAQKRSKYIAKAFISSFIAITVGAGITLAVLILSGSIEFIPMQV 132

Query: 125 IPVGGMIISNSMVAIGLCYKQLLSEFRSKQEEVETKLALGADILPASIDIIRDVIKTMV 184
      IP+ GMI N+MVA+GLCY L S+Q++++ KL+LGA AS +IRD I+ ++
40 Sbjct: 133 IPIAGMLAGNAMVAVGLCYNNLGRVISEQQQIQEKLKSLGATPKQASAILIRDSIRAALI 192

Query: 185 PTIDSAKTGIVSLPGMMTGLILAGTSPIQAVKYQMMVTFMILLATTSIASFVATYLYAYKI 244
      PT+DSAKT+G+VSLPGMM+GLI AG P++A+KYQ+MVTFMLL+T S+++ +A YL Y+
45 Sbjct: 193 PTVDSAKTVGLVSLPGMMSGLIFAGIDPVKAIKYQIMVTFMLLSTASLSTIIACYLT YRK 252

Query: 245 FFNNRKQLVVT 256
      F+N+R QLVVT+
50 Sbjct: 253 FYNRRQLVVTQ 264

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 397

A DNA sequence (GBSx0431) was identified in *S.galactiae* <SEQ ID 1289> which encodes the amino acid sequence <SEQ ID 1290>. This protein is predicted to be SUGAR TRANSPORT ATP-BINDING PROTEIN. (b0490). Analysis of this protein sequence reveals the following:

5 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1903 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GF:AAC73592 GB:AE000155 putative ATP-binding component of a
 transport system [Escherichia coli K12]
 Identities = 95/202 (47%), Positives = 142/202 (70%), Gaps = 2/202 (0%)

20 Query: 4 LTFKHVDFKTDKLVNDINFAIDEGDFVSI VGPSSGSGKSTVLKSLASGLMSPTAGHIFD 63
 L ++V + D +LN+INF++ G+F I GPSG GKST+LK+ + L+SPT+G + F+
 Sbjct: 8 LQLQNVGYLAGDAKILNNINFSLRAGEFKLITGPSGCGKSTLLKIVASLISPTSGTLLFE 67

25 Query: 64 GKDLNQLLEPIESRKMI SYCFQTPHLFGNTVEDNISFPYHIRHEKVDYRRVDDLQRFEMD 123
 G+D++ L+P R+ +SYC QTP LFG+TV DN+ FP+ IR+ + D D +RF +
 Sbjct: 68 GEDVSTLKPETIRQQVSYCAQTPTLFGDTVYDNLIFPWQIRNRQDPAPFLDFLERFALP 127

30 Query: 124 QSYLKQDVKKLSGGGEKQRIALIRQLLFEPKVL LLDDEVTSALDNHNKAIVEKVI-KSLHDK 182
 S L +++ +LSGGEKQRI+LIR L F PKVLLLDE+TSALD NK V ++I + + ++
 Sbjct: 128 DSILTKNIAELSGGEKQRI SLIRNLQFMPKVLLLDEITSALDESNNKHNVNEMIHRYVREQ 187

30 Query: 183 GITILWITHDEEQSRRFANKVL 204
 I +LW+THD+++ A+KV+
 Sbjct: 188 NIAVLWVTHDKDEINH-ADKVI 208

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1291> which encodes the amino acid
 sequence <SEQ ID 1292>. Analysis of this protein sequence reveals the following:

 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2053 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

45 Identities = 73/214 (34%), Positives = 133/214 (62%), Gaps = 9/214 (4%)

50 Query: 4 LTFKHVD--FKTDDKLVNDINFAIDEGDFVSI VGPSSGSGKSTVLKSLASGLMSPTAGHIF 61
 +TF +V F+ VL +INF ++EG F +++G SGSGKST+L + +GL+ ++G I+
 Sbjct: 6 ITFNNVSKTFEDSGTQVLKNINFDLEEGKFYTL LGASGSGKSTILNIMAGLLDASSGDIY 65

55 Query: 62 FDGKDLNQLLEPIESRKMI SYCFQTPHLFGN-TVEDNISFPYHIR--HEKVDYRRVDDLQ 118
 DG+ +N L PI R I FQ LF + TV +N++F ++ +K +RV + +
 Sbjct: 66 LDGERINDL-PINKRD-IHTVFQNYALFPHMTVPENVAFALKLKKVVDKKEIAKRVKETLK 123

60 Query: 119 RFEMDQSYLKQDVKKLSGGGEKQRIALIRQLLFEPKVL LLDDEVTSALDNHNKAIVEKVIKS 178
 ++ + + + ++KLSGG++QR+A+ R ++ +P+V+LLDE SALD + ++ ++
 Sbjct: 124 MVQL-EGFENRSIQKLSGGQRQVARIARAIINQPRVLLDEPLSALDLKLRTEMQYELRE 182

60 Query: 179 LHDK-GITILWITHDEEQSRRFANKVLKVNGSI 211
 L + GIT +++THD+E++ ++ + + G I
 Sbjct: 183 LQQR LGITFVFVTHDQEEALAMSDWIFVMNEGEI 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 398

- 5 A DNA sequence (GBSx0432) was identified in *S.agalactiae* <SEQ ID 1293> which encodes the amino acid sequence <SEQ ID 1294>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0658(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 399

- 20 A DNA sequence (GBSx0434) was identified in *S.agalactiae* <SEQ ID 1295> which encodes the amino acid sequence <SEQ ID 1296>. This protein is predicted to be deda protein (dedA). Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

- 25 INTEGRAL Likelihood = -12.05 Transmembrane 186 - 202 (178 - 208)
INTEGRAL Likelihood = -8.81 Transmembrane 65 - 81 (61 - 89)
INTEGRAL Likelihood = -7.54 Transmembrane 26 - 42 (24 - 47)
INTEGRAL Likelihood = -0.37 Transmembrane 152 - 168 (152 - 168)

- 30 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC75377 GB:AE000320 orf, hypothetical protein [Escherichia coli K12]
Identities = 91/211 (43%), Positives = 131/211 (61%), Gaps = 7/211 (3%)

- 40 Query: 2 FLIDFILHIDTHIYAMANTVGNWYLLFLVIFVETGAVIFPFLPGDSLLFAAGALAANP 61
FLIDFILHID H+ + G W Y +LFL++F ETG V+ PFLPGDSLLF AGALA+
Sbjct: 6 FLIDFILHIDVHLAELVAEYGVVYAILFLILFCETGLVVTPLPGDSLLFVAGALASLE 65

- Query: 62 KMSFNIVTFLIIFFFIAAFIGDSCNFIIGRTFGYRFIKHP---FFRRFIKEKNIRDALYF 118
N+ +++ IAA +GD+ N+ IGR FG + +P FRR +K ++
45 Sbjct: 66 TNDLNVHMMVVLMLIAAIVGDAVNYITIGRLFGEKLFSPNSKIFRRSYLDK---THQFY 121

- Query: 119 EKKGTAAIILGRYIPIIRTFFVPFVAGISQLPPKVFIFKRAFIAALSWSVIATGSGFLFGNI 178
EK G IIL R++PI+RTF PFVAG+ + + F I AL W ++ T +G+ FG I
Sbjct: 122 EKHGGKTIILARFVPIVRTFAPFVAGMGHMSYRFAAYNVIGALLWVLLFTYAGYFFGTI 181

- 50 Query: 179 PFVKQHFSLIILGIVFVTLIPVLISGVKSYR 209
P V+ + L+I+GI+ V+++P +I ++ R
Sbjct: 182 PMVQDNLKLIVGLIIVSILPGVIEIRHKR 212

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 400

- 5 A DNA sequence (GBSx0435) was identified in *S.agalactiae* <SEQ ID 1297> which encodes the amino acid sequence <SEQ ID 1298>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3100 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 401

- 20 A DNA sequence (GBSx0436) was identified in *S.agalactiae* <SEQ ID 1299> which encodes the amino acid sequence <SEQ ID 1300>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3990 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9323> which encodes amino acid sequence <SEQ ID 9324> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
Identities = 87/157 (55%), Positives = 110/157 (69%), Gaps = 1/157 (0%)

Query: 1 MLDRTIRQYQNRDRDTLPDANWKPLGWHQVAT-NDHYGHAVDKGHLIAYALAGNFKGWDA 59
+L + RQY+NR++T +W P GWHQV Y HAVD+GHL+ YAL G G+DA
Sbjct: 116 LLSKATROQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDA 175

Query: 60 SVSNPQNVTQTAAHSNQSNOQINRGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPP 119
S SNP+N+ QTA +NQ+ + + GQNYYES VRKA+DQNKRVRYRV Y ++ DLVP
Sbjct: 176 STSNPKNIQVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPS 235

Query: 120 AMHLEAKSQDGTLEFNVAIPNTQASYTMDYATGEITL 156
A +EAKS DG LEFNV +PN Q +DY TGE+T+
Sbjct: 236 ASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTV 272

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1301> which encodes the amino acid sequence <SEQ ID 1302>. Analysis of this protein sequence reveals the following:

-505-

Possible site: 42
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 89/135 (65%), Positives = 104/135 (76%), Gaps = 1/135 (0%)

 Query: 25 SPAGWHRLHHLKGSYDHA VDRGHLLGYALVGGLKGFDA STGNPDNIATQLSWANQANKPY 84
 +P GWH++ +LKGSY HAVDRGHLLGYAL+GGL GFDAST NP NIA Q +WANQA Y
15 Sbjct: 138 TPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDA STSNPKNIAVQTAWANQAQAEY 197

 Query: 85 LTGQNYEGLVRRALDKGHRVRYRVTLTY-DGDNLLASGSHLEAKSSDDSLTFNVFVFNV 143
 TGQNYE VR+ALD+ RVRVRYTL Y ++L+ S S +EAKSSD L FNV VPNV
20 Sbjct: 198 STGQNYYESKVRKALDQNKRVRYRVTLTYASNEDLVPSASQIEAKSSDGELEFNVLVPNV 257

 Query: 144 QAGLTADYRTGQIAI 158
 Q GL DYRTG++ +
 Sbjct: 258 QKGLQLDYRTGEVTV 272

25 An alignment of the GAS and GBS proteins is shown below:

 Identities = 73/135 (54%), Positives = 92/135 (68%), Gaps = 2/135 (1%)

 Query: 24 PLGWHQVA-TNDHYGHA VDKGHLLIAYALAGNFKGWDASVSNPQNVVTQTAAHSNQSNOKIN 82
 P GWH++ Y HAVD+GHL+ YAL G KG+DAS NP N+ TQ + +NQ+N+
30 Sbjct: 26 PAGWHRLHHLKGSYDHA VDRGHLLGYALVGGLKGFDA STGNPDNIATQLSWANQANKPYL 85

 Query: 83 RGQNYYESLVRKA VDKQNKRVRYRVTPLYRNDTDLVFPFAMHLEAKSQDGTLEFNVAIPNTQ 142
 GQNYE LVR+A+D+ RVRVRYTL Y D +L+ HLEAKS D +L FNV +PN Q
35 Sbjct: 86 TGQNYEGLVRRALDKGHRVRYRVTLTYDGD-NLLASGSHLEAKSSDDSLTFNVFVFNVQ 144

 Query: 143 ASYTM DYATGEITLN 157
 A T DY TG+I +N
 Sbjct: 145 AGLTADYRTGQIAIN 159

40 SEQ ID 9324 (GBS656) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 186 (lane 10; MW 57kDa).

GBS656-GST was purified as shown in Figure 236, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 **Example 402**

A DNA sequence (GBSx0437) was identified in *S.agalactiae* <SEQ ID 1303> which encodes the amino acid sequence <SEQ ID 1304>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

50 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
55

-506-

A related GBS nucleic acid sequence <SEQ ID 9321> which encodes amino acid sequence <SEQ ID 9322> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1305> which encodes the amino acid sequence <SEQ ID 1306>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5350(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 24/73 (32%), Positives = 37/73 (49%), Gaps = 2/73 (2%)

Query: 1 MFYMKLIANRLSLAATIIVNEANANSPFGIIHSDKAENVEWPDFETQFPDLFNSPKKEESP 60
+ YMKLA L TI+ E + SPF I+H+D A N++ E N +++P
Sbjct: 80 ILYMKLAKENHLPVTIITETHMTSPFAFILHTDHAINLKETRLEVILKQTKNDQLSKQTP 139

Query: 61 K--KSLWQHFFSQ 71
+ KS W+ F +
Sbjct: 140 EKTKEFWKRFLKK 152

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 403

A DNA sequence (GBSx0438) was identified in *S.galactiae* <SEQ ID 1307> which encodes the amino acid sequence <SEQ ID 1308>. This protein is predicted to be Isopentenyl-diphosphate delta-isomerase.

Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG20030 GB:AE005083 isopentenyl pyrophosphate isomerase; Idi
[Halobacterium sp. NRC-1]
Identities = 24/77 (31%), Positives = 40/77 (51%)

Query: 14 TGLTLNRDQNIPOQLFHLVVDVILFHEDGDVLMKMRHPPKKAFFPAYFEATAGGSALKGEN 73
TGL D + G+ H +LF EDG VL+ +R +K+ + +++ T ++G++
Sbjct: 42 TGLANRLDAHTGDGVRHRAFTCLLFDEIDGRVLLAQRADRKRLWDTHWDGTVASHPIEQS 101

Query: 74 AKQAILRELKEETGIVP 90
A + L EE GI P

Sbjct: 102 QVDATRQRLAELGIEP 118

No corresponding DNA sequence was identified in *S.pyogenes*.

-507-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 404

5 A DNA sequence (GBSx0439) was identified in *S.agalactiae* <SEQ ID 1309> which encodes the amino acid sequence <SEQ ID 1310>. This protein is predicted to be phosphoserine phosphatase (serB). Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0613(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB50876 GB:AL096844 putative phosphoserine phosphatase
[Streptomyces coelicolor A3(2)]
Identities = 96/193 (49%), Positives = 132/193 (67%)

20 Query: 5 LLVMDVDSTLIMEEAIIDLIAIEAGVGKQVAALTDAAAMRGELDFEEALKKRVALLKGLFVT 64
L+VMDVDSTLI +E I+L A AG +VA +T AAMRGELDPE++L RVALL GL +
Sbjct: 183 LVVMDVDSTLIQDEVIELFAAHAGCEDEVAEVTAAAMRGELDFEQLHARVALLAGLDAS 242

25 Query: 65 ILTDILSSIHFTPGAYELIKECHKRQMKVGLVSGGFHETIDILAKQLQVDYVKANRLGVK 124
++ + + + TPGA LI+ + +VG+VSGGF + D L +QL +D+ +AN L +
Sbjct: 243 VVDKVRAEVRLTPGARTLIRTLKRLGYQVGVSFGFTQVTDALQEQGLDFAQANTLEIV 302

Query: 125 GGFLTGEVEGEIVTKEVKKIKLEWASENHLDSLQTIAMGDGANDLPMIKSAGVGIAFCA 184
G LTG V GEIV + K L+ +A+ + LSQT+A+GDGANDL M+ +AG+C+AF A
30 Sbjct: 303 DGRLTGRVTGEIVDRAGKARLRRFAAAAGVPLSQTVAIIGDGANDLMLNAAGLGVAFNA 362

Query: 185 KPIVREEAAYQIN 197
KP+VRE A +N
35 Sbjct: 363 KPVVREAAHTAVN 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 405

40 A DNA sequence (GBSx0440) was identified in *S.agalactiae* <SEQ ID 1311> which encodes the amino acid sequence <SEQ ID 1312>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-17.88 Transmembrane 5 - 21 (1 - 29)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.8153(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 122/553 (22%), Positives = 265/553 (47%), Gaps = 12/553 (2%)

-508-

Query: 7 LLLVAIVLLVIIAYVVGVIIRKNDTLIANLETRKQELVDLPVQEEIEQVKLLHLIGSQ 66
 +++ +++L + +V G + RK + LE K +++ P+ +EI +VK L + G+++
 Sbjct: 3 IVVFSLLVLVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIDBIGVKGLTMSGETE 62

5 Query: 67 STPREWNQKWTDLSTNSFKDIDFHLVEAENLNSFNFRVRAKHEIDNVDSQLTIIIEEDIVS 126
 F W W D+ +++ L + E+ + + F +AK +D ++ +L IEE +
 Sbjct: 63 EKFEVWRSDWDDIVGVILPNVEEQLFVDFANKYRFPQAKALLDTIEQRLHSIEEQLKI 122

10 Query: 127 IREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNIEAEFSH 186
 + + ++VL + EE+N + +L + L K + + ++ +++L+
 Sbjct: 123 MVDIDIQLVLVQSEEQNRTEIGSVRELQKLIKELATRRGSLSSSAKFDEKLEKANELLOA 182

15 Query: 187 FVTINSTGDPRIEASEVLNKAEEHTIALGQITEQIPAIKLEDDFPDQLDDLETGYRRL 246
 F G+ I+ASEVL +A+E + + + +P + +L+ + P +L +L+ G R +
 Sbjct: 183 FDERTEKGNITQASEVLEAKELGQIEHLLKIVPGLFVELQTNIPAEITNLKNGLRDM 242

20 Query: 247 EENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRDENEHIQEKIDKLYDIFEREIA 306
 E + I+ + + + E + L L+ + +E I+E ++++++ E+E+
 Sbjct: 243 EAGFFLETFATDSQMERLEEKRVLEQLTVLECNMEEETINFIEESMEQMFELLEKEVE 302

25 Query: 307 AYKVAHKDSKIIPQFLAHAKSNNEQLCH---EIKRLSAKYILNENESLSLRSFTNDLEEI 363
 A ++ + ++P E+L H E + + Y L E E + + +L+E+
 Sbjct: 303 A---KNEITILLPNLREDLTKTEBKLTHLKEETBSVQLSYRLAEEELVFQKLGKELKEL 359

30 Query: 424 DKYVKNLHMIKRFMEKRNLPGLIPQDFLSTFTTSSQIEALINELSRGRIDIEAVSRLNDV 483
 + KL KR ++K N+PG+P+ L ++ I +LS +++ V+ L D
 Sbjct: 420 KQLKEKILEDKRLVQKSNIPGLPETLHRLLEDGEQKLAQAIKLSDVPLEMGRVTALVDE 479

35 Query: 484 TTNIAIANLEQATYLVVDATLTLEQLQYSNRYRSFEQNVQKSFEQALYLFVEHNYKASF 543
 I + ++ A L E ++QY NRYRS V+K A LF +
 Sbjct: 480 AQLIHENSSILHETIEKARLAEHVIQYGNRYRSRAEVKKRLSNAEELFRA-----FEY 534

Query: 544 DE-ISALETVEP 555
 DE I A++ +EP
 Sbjct: 535 DEAIEMAVQAIEP 547

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1313> which encodes the amino acid sequence <SEQ ID 1314>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 45 INTEGRAL Likelihood = -18.04 Transmembrane 5 - 21 (1 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.8217(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 131/555 (23%), Positives = 269/555 (47%), Gaps = 16/555 (2%)

55 Query: 7 LLLVAIVLLVIIAYLVGVIIRKNDLITSLERKQALFALPVNDEIEEVKSLHLIGSQ 66
 +++ +++L + ++ G + RK + LE+ K + P+ DEI +VK L + G+++
 Sbjct: 3 IVVFSLLVLVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIDBIGVKGLTMSGETE 62

60 Query: 67 TSFREWQKWDLTIVNSFADIENHIFEAEENLNTFNFIRAKHEINSVESQLNLVEEDIAS 126
 F W W D+ ++E +F+ E+ + + F +AK +++++E +L+ +EE +
 Sbjct: 63 EKFEVWRSDWDDIVGVILPNVEEQLFVDFANKYRFPQAKALLDTIEQRLHSIEEQLKI 122

65 Query: 127 IREALNILEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEDKQMKNIETEF 186
 + + + +L + EE+N + +L +KL + S+ D++++

Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQKQLIKEAITRRGSLSSSAKVFEKLEKANELLQA 182

Query: 187 FVALNSSGDPVEASEVLDRAEHTIALGQITEQIPAIKLEDDFPDQLDDLETGYRRL 246
F G+ ++ASEVL+ A+E + + + +P + +L+ + P +L +L+ G R +

5 Sbjct: 183 FDERTEKGNIIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPAEITNLKNGLRDME 242

Query: 247 EENYHFPEKNIEARFQEIRESIRANSSELVTLLDRAREENTHIQERIDSLYEVFEREIA 306
E + I+++ + + E +L L+ + EE I+E ++ ++E+ E+E

10 Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNMGEEINFIEESMEQMFELLEKE-- 300

Query: 307 AYKVAKN--SKMLPRYLEHVKNRNEQ---LKDEIARLSRKYILSETESLTVKAFEKDIK 361
V AKN + +LP E + + E+ LK+E + Y L+E E + + K++K

Sbjct: 301 ---VEAKNEITILLPNLREDLTKEEKLTHLKEETESVQLSYRLAEELVFPQKLGKELK 357

15 Query: 362 EIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQARH 421
E+ + E Q++ FS ++ E + LT ++ ++ + K E +A+

Sbjct: 358 ELRQQLQVIDEVTEEQKQTFSSVSRMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKE 417

20 Query: 422 NLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRLS 481
L +L KR ++K ++PG+P+ L +L + +LS + + V+ L

Sbjct: 418 ELKQLKEKLEDKRLVQKSNIPGLPETLLHRLLEDGEQKLAQAIKLSDVPLEMGRVTALV 477

Query: 482 EVATVAIANLEDITYQVQVQNTATLFEQLLQYSNRYRSFEAGVQSSFHALRLFEVENDYQA 541
+ A I + ++ ++ A L E ++QY NRYRS A V+ +A LF

25 Sbjct: 478 DEAQGLIHENSSILHETIEKARLAHVQYGNRYRSRAEVKKRLSNAEELFRA-----F 532

Query: 542 SFDE-ISYALETVEP 555
+DE I A++ +EP

30 Sbjct: 533 EYDEAIEMAVQAIEP 547

An alignment of the GAS and GBS proteins is shown below:

Identities = 429/574 (74%), Positives = 503/574 (86%)

35 Query: 1 MSSGIILLVAIVLLVIAIYVGVVIRKNDTLIANLETRKQELVDLPVQEEIEQVKLLH 60
MSSGIILL+VAIVLLVIAIY+VGV+IRKND+LI +LE RKQ L LPV +EIE+VK LH
Sbjct: 1 MSSGIILLVAIVLLVIAIYLVGVVIRKNDLITSLEERKQALFALPVNDEIEEVKSLH 60

40 Query: 61 LIGQSQSTFREWQKWTDLSTNSFKDIDFHLVEAENLNDSENFVRAKHEIDNVDSQLTII 120
LIGQSQ++FREWQKW DL+ NSF DI+ H+ EAENLND+FNH+RAKHEI++V+SQL ++
Sbjct: 61 LIGQSQSTFREWQKWVDLTVNSFADIEHIFEAENLNDTFNFIRAKHEINSVESQLNLV 120

45 Query: 121 ERDIVSIREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTMPEIEKQLKNI 180
EBDI SIREAL +LKEQEEKNSARVTHALDLYE LQ SISE EDN+G+TMPEI+KQ+KNI
Sbjct: 121 ERDIASIREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNI 180

50 Query: 181 EAEFSHFVTLNSTGDPTEASEVLNKAEEHTIALGQITEQIPAIKLEDDFPDQLDDLET 240
E EFS FV LNS+GDP+EASEVL++AEHTIALGQITEQIPAIKLEDDFPDQLDDLET
Sbjct: 181 ETEFSQFVALNSSGDPVEASEVLDRAEHTIALGQITEQIPAIKLEDDFPDQLDDLET 240

55 Query: 241 GYRRLLEENYHFPEKDIEQRQFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDI 300
GYRRLLEENYHFPEK+IE RFQE+RE+IR+NS LV+LDLDRAR+EN HIQE+ID LY++
Sbjct: 241 GYRRLLEENYHFPEKNIEARFQEIRESIRANSSELVTLLDRAREENTHIQERIDSLYEV 300

60 Query: 301 FERIEAAYKVAHKDSKIIPQFLAHAKSNNEQLGHEIKRLSAKYILNENESLSRSFTNDL 360
FERIEAAYKVA K+SK++P++L H K NNEQL EI RLS KYIL+E ESL++++F D+
Sbjct: 301 FERIEAAYKVAANKSKMLPRYLEHVKNRNEQLKDEIARLSRKYILSETESLTVKAFEKDI 360

65 Query: 361 EBIETKVLPSVENFGQEASPYTHLQILFERTLKTTLTVEENQMEVFEAVKTIESVETRAR 420
+EIE L E FG + P++ LQ+ FER++KTLT VE QM+VF AVK IE +E++AR
Sbjct: 361 KEIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQAR 420

Query: 421 QNMDKYVNKLHMIKRFMEKRNLPGLPQDFLSTFTTSSQIEALINELSRGRIDIEAVSRL 480
N+D YV +LHMIKR+MEKR+LPGIPQDFLS FTTTSSQ+EAL++ELSRGRI+IEAVSRL
Sbjct: 421 HNLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRL 480

Query: 481 NDVTINAIAANLEQATYLVVQDATLFEQLLQYSNRYRSFEQNVQKSFEQALYLFVEVHNYK 540
++V T AIANLE TY VVQ+ATLFEQLLQYSNRYRSFE VQ SFE AL LFEVE++Y+

-510-

Sbjct: 481 SEVATVAIANLEDLTQVQVQATLLEQLLQYSNRYRSFEAGVQSSFEHALRLFEVENDYQ 540

Query: 541 ASFDEISYALETVEPGVTDTRFVTSYEKTRERIRF 574

ASFDEISYALETVEPGVTDTRFV SYEKT+E IRF

5 Sbjct: 541 ASFDEISYALETVEPGVTDTRFVNSYEKTRERIRF 574

SEQ ID 1312 (GBS642) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 406

A DNA sequence (GBSx0441) was identified in *S.agalactiae* <SEQ ID 1315> which encodes the amino acid sequence <SEQ ID 1316>. Analysis of this protein sequence reveals the following:

15 Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.2471(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9671> which encodes amino acid sequence <SEQ ID 9672> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:CAA91553 GB:Z67740 DNA gyrase [Streptococcus pneumoniae]
Identities = 574/650 (88%), Positives = 618/650 (94%), Gaps = 2/650 (0%)

Query: 1 MTEETKNMEQRAQEYDASQIQVLEGLEAVMRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
MTEE KN++ AQ+YDASQIQVLEGLEAVMRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA

30 Sbjct: 1 MTEBIKNLQ--AQDYDASQIQVLEGLEAVMRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 58

Query: 61 LAGFAGHIKVYIEPDNSITVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
LAGFA HI+V+IEPD+STITVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV

35 Sbjct: 59 LAGFASHIQVFIEPDNSITVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 118

Query: 121 SGGHLGVGSSVNALSTQLDVVKYKNGKVHYQEYQRGVVNDLEIIGDITDLSGTTVHFTP 180
SGGLHGVGSSVNALSTQLDV V+KNGK+HYQEY+RG VV DLE++GDTD +GTTVHFTP

40 Sbjct: 119 SGGHLGVGSSVNALSTQLDVHVKNGKIHVQEYRRGHVVADLEVVDGTDRTGTTVHFTP 178

Query: 181 DPEIFTETTVDFDKLAKRIQELAFNLRLGISIDKREGQVEVEKEYHYEGGIGSYVEFI 240
DPEIFTETT+FDFDKL KRIQELAFNLRL+ISI+DKR+G E K YHYEGGI SYVE+I

45 Sbjct: 179 DPEIFTETTFDFDKLNKRIQELAFNLRLQISITDKRQGLEQTKHYHYEGGIASVEYI 238

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFR TAL 300
NENK+VIF+ PIYTDGE+D I+VEVAMQYTTGY E VMSFANNIHTHEGGTHEQGFR TAL

50 Sbjct: 239 NENKDVIFDTPYITDGEDDITVEVAMQYTTGYHENVMSFANNIHTHEGGTHEQGFR TAL 298

Query: 301 TRVINDYAKGNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
TRVINDYA+KNK+LK+NEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT

55 Sbjct: 299 TRVINDYARKNKLKDNEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 358

Query: 361 NRLFSEAFNRLLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGLAD 420
NRLFSEAF+ FL+ENPQ+AK+IVEKGILA+KAR+AAKAREVTRKKSGLEISNLPGLAD

Sbjct: 359 NRLFSEAFSDFLMENPQIAKRIVEKGILAAKARVAAKAREVTRKKSGLEISNLPGLAD 418

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
CSSNN ELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKA+MDKILANEEIRS

Sbjct: 419 CSSNPAETELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKASMDKILANEEIRS 478

-511-

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540
 LFTAMGTGFGA+FDVSK RYQKLV+MTDADVDGAHIRTLLLTLYR+M+P+LEAGYVYIA
 Sbjct: 479 LFTAMGTGFGAEFDVSKARYQKLVIMTDADVDGAHIRTLLLTLYRYMKPILVLEAGYVYIA 538

Query: 541 QPPIYGKVGSEIKAYIQPGVNQEBELRQALDTSYSSGRSKPTVQRYKGLGEMDDHQLWET 600
 QPPIYGKVGSEIK YIQPG +QE +L++AL YS GR+KPT+QRYKGLGEMDDHQLWET
 Sbjct: 539 QPPIYGKVGSEIKEYIQPGADQBIKQLEALARYSEGRTKPTIQRKGLGEMDDHQLWET 598

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
 TMDPE+RLMARVSVDDAAEADKIFDMLMGDRVEPRREFIE NAVYS LD+
 Sbjct: 599 TMDPEHRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSTLDV 648

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1317> which encodes the amino acid
 sequence <SEQ ID 1318>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 584/650 (89%), Positives = 618/650 (94%)

Query: 1 MTEETKNMEQRAQEYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
 M EE K+ E++ QEYDASQIQVLEGLEAVRMRPGMYIGST+KEGLHHLVWEIVDNSIDEA
 Sbjct: 1 MIEENKHFEEKMQQEYDASQIQVLEGLEAVRMRPGMYIGSTAKEGLHHLVWEIVDNSIDEA 60

Query: 61 LAGFAGHIKVYIEPDNSITVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
 LAGFA HIKV+IE DNSITVDDGRGIPVDIQ KTGRPAVETVFTVLHAGGKFGGGGYKV
 Sbjct: 61 LAGFASHIKVFIEADNSITVDDGRGIPVDIQAKTGRPAVETVFTVLHAGGKFGGGGYKV 120

Query: 121 SGGLHGVGSSVVALSTQLDVKVYKNGKVHYQEYQRCVVNDLEIIGDSDLGTTVHFTP 180
 SGGLHGVGSSVVALSTQLDV+VYKNG++HYQE++RG VV DLE+IG TD++GTTVHFTP
 Sbjct: 121 SGGLHGVGSSVVALSTQLDVRVYKNGQIHYQEYFRCGAVVADLEVIGTTDVIGTTVHFTP 180

Query: 181 DPEIFTETTVDFDKLAKRIQELAFNLRLGISISDKREGQEVEKEYHYEGGIGSYVEFI 240
 DPEIFTETT FD+ LAKRIQELAFNLRLGL+ISI+DKR G E + YEGGIGSYVEF+
 Sbjct: 181 DPEIFTETTFQDYSVLAKRIQELAFNLRLGLKISITDKRSGMEQEEHFLYEGGIGSYVEFL 240

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHERQGFRTAL 300
 N+ K+VIFE PIYTDGEL+GI+VEVAMQYTT YQETVMSFANNIHTHEGGTHERQGF AL
 Sbjct: 241 NDKKDVIFETPIYTDGELGIEAVEVAMQYTTSYQETVMSFANNIHTHEGGTHERQGFRAAL 300

Query: 301 TRVINDYAKKKNILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
 TRVINDYAKKKNILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
 Sbjct: 301 TRVINDYAKKKNILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360

Query: 361 NRLFSEAFNRFLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGLAD 420
 NRLFSEAF RFLLENPQVA+KIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGLAD
 Sbjct: 361 NRLFSEAFQRFLENPQVARKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGLAD 420

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
 CSSN+A NELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS
 Sbjct: 421 CSSNDANQNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540
 LFTAMGTGFGADFDVSK RYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA
 Sbjct: 481 LFTAMGTGFGADFDVSKARYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540

Query: 541 QPPIYGKVGSEIKAYIQPGVNQEBELRQALDTSYSSGRSKPTVQRYKGLGEMDDHQLWET 600
 QPPIYGKVGSEIK YIQPG++QE++L+ AL+ YS GRSKPTVQRYKGLGEMDDHQLWET
 Sbjct: 541 QPPIYGKVGSEIKEYIQPGIDQEDQLKTALEKYSIGRSKPTVQRYKGLGEMDDHQLWET 600

-512-

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
 TMDPENRLMARV+VDDAAEADK+FDMLMGDRVEPRR+FIE NAVYS LDI
 Sbjct: 601 TMDPENRLMARVTVDDAAEADKVFDMLMGDRVEPRRDFIEANAVYSTLDI 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 407

A DNA sequence (GBSx0442) was identified in *S.agalactiae* <SEQ ID 1319> which encodes the amino acid sequence <SEQ ID 1320>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3186(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
 Identities = 82/142 (57%), Positives = 105/142 (73%)

Query: 45 LKESTADAIAYFIPEEADFLKEYKANEAKVLETPILFQGAKEALLAKIQRQGSRNFLVSHR 104
 LK ST AI F P +FL++YK NEA+ LE PILF+G +LL I QG R+FLVSHR

Sbjct: 2 LKVSTPFAIETFPAPNLNLFLEKYKENEARELEHPILFEGVSDLEEDILNQGGRHFLVSHR 61

Query: 105 DNQVIVILEKTEIIDYFTEVVTTADNGFSRKPSPEMMLYLKEYQIDNCLVIGDRDRIDKQA 164
 ++QV+ ILEKT I YFTEVVTT+ +GF RKP+PESMLYL+EKYQI + LVIGDR ID +A

Sbjct: 62 NDQVLEILEKTSIAAYFTEVVTTSSSGFKRKNPESMLYLREKYQISSGLVIGDRPIDIEA 121

Query: 165 GESAGFDTLVLDGSKSLMEIIE 186
 G++AG DT L +L ++++

Sbjct: 122 GQAAGLDTHLFTSIVNLRQVLD 143

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1321> which encodes the amino acid sequence <SEQ ID 1322>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2472(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 122/185 (65%), Positives = 145/185 (77%)

Query: 1 MNYHDIWDLGGTLLDNYESSSTRAFVETLKEFGYQADHDSVYQKLKESTADAIAYFIPEE 60
 MNY DYIWDLGGTLLDNYE ST+AFV+TL F DHD+VYQKLKESTA A+A F P E

Sbjct: 4 MNYQDIWDLGGTLLDNYELSTQAFVQTLAFTSLPGDHDAVYQKLKESTAIAMFAPNE 63

Query: 61 ADFLKEYKANEAKVLETPILFQGAKEALLAKIQRQGSRNFLVSHRDNQVIVILEKTEIIDY 120
 +FL Y+ EA L PI GAKE+L KI GSRNFL+SHRD QV +LE+ ++ Y

Sbjct: 64 PEFLHVYRLREADKLAQPIWCLGAKEILGKIATSGSRNFLISHRDCCQVNLLEQAGLLIY 123

Query: 121 FTEVVTTADNGFSRKPSPEMMLYLKEYQIDNCLVIGDRDRIDKQAGESAGFDTLVLDGSKS 180
 FTEVVIA NGF+RKP+PES+ YLKEY I++ LVIGDR IDKQAG++AGF+TLVLDG K+

Sbjct: 124 FTEVITASNGFARKPNPESLFYLKEYDINSGLVIGDRDLIDKQAGQAGFNTLLVLDGRKN 183

-513-

Query: 181 LMEII 185
 L+EI+
 Sbjct: 184 LLEIV 188

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 408

A DNA sequence (GBSx0443) was identified in *S.agalactiae* <SEQ ID 1323> which encodes the amino acid sequence <SEQ ID 1324>. This protein is predicted to be stage V sporulation protein E (rodA).

- 10 Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -11.15	Transmembrane	206 - 222 (177 - 226)
INTEGRAL	Likelihood = -10.14	Transmembrane	58 - 74 (50 - 82)
INTEGRAL	Likelihood = -9.34	Transmembrane	182 - 198 (177 - 205)
INTEGRAL	Likelihood = -8.55	Transmembrane	158 - 174 (156 - 177)
INTEGRAL	Likelihood = -8.12	Transmembrane	300 - 316 (299 - 324)
INTEGRAL	Likelihood = -2.66	Transmembrane	86 - 102 (83 - 102)
INTEGRAL	Likelihood = -2.34	Transmembrane	338 - 354 (338 - 357)

----- Final Results -----
 bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9669> which encodes amino acid sequence <SEQ ID 9670> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15838 GB:Z99123 alternate gene name: ipa-42d~similar to
 cell-division protein [Bacillus subtilis]
 Identities = 142/392 (36%), Positives = 237/392 (60%), Gaps = 23/392 (5%)

Query: 10 QKSNFYKQIDYAVVIVPVFFLLMIGLASIYVA-TMNDYPSNIYIAMFQQVSWIIMGCIIA 68
 Q+S +++G D + VFF+ I + SIY A Y + +I QQ+ + ++G +
 Sbjct: 7 QQSPFYQG--DLIFIFGVFFI--ISVVSIIAAGQFGQYGNWDWI---QQIVFYLLGAVAL 59

Query: 69 FVVMLEFSTFLWKATPYLYALGLTLMVLPLIFYSPLFAAT--GAKNWTIGSVTLFQPS 126
 V++ F E L K + Y++ +G+ +++ I SP+ A GAK+W IG +T+ QPS
 Sbjct: 60 TVLLYFDLEQLEKLSLYIFILGILSLIILKI--SPESIAPIVKGAKSWFRIGRITI-QPS 116

Query: 127 EFMKISYILMLSRITVSFHQNRKTFQDDWKL--GLFGLVTLPMILLMLQKDLGTALVF 185
 EFMK+ I+ML+ + + K +T +DD LL + G+ +PV ++LM +D GTA +
 Sbjct: 117 EFMKVGLIMMLASVIGKANPKGVRTLDDIHLKLLKIAGVAVIPVGLIIM--QDAGTAGIC 174

Query: 186 LAILSGLILLSGISWIIILPILSTIVLFIAFLMIFISPNGKEWFYNLGMPTYQINRLSA 245
 + I+ ++ +SGI+W +I I + +L I+ L++ I N + ++G+ YQI R+++
 Sbjct: 175 MFIVLVVMVMSGINWKLIAITAGSGILLISLILLVMI--NFPDVAKSVGIQDYQIKRVT 232

Query: 246 WIDPFSFAD---SIAYQQTQGMVSGISGGVTGKGFNILELSVPVRESDMIPTVIAENFGF 302
 W+ + + ++Q Q +++IGSGG+ G G + L++ VP +D IF++I E+FGF
 Sbjct: 233 WVSASNETQEDSNDSQVDQAIMAIGSGGILGNGISNLKVYPSTTDFIFSIIIGESFGF 292

Query: 303 IGSAIVLGLYLIIIYRMLRIT--IESNNQFYTFISTGFIMMIVFHVFNENIGAAGVILPLT 360
 IG AIV+ ++ +IYR++ + I N+F +F G+ +IV H F+NIG +GI+P+T
 Sbjct: 293 IGCAIVVIMFFFLIYRLVVLIDKIHFPNRFASFVFCVGTALIVIHFTQNIQNMNIGIMPVT 352

Query: 361 GIPLPFISQGGSSLLSNLIGLVLMSYQNT 392
 GIPL F+S GGSS LS LIG G+V + S Q T
 Sbjct: 353 GIPLLFVSYGGSSTLSTLIGFIVYNASVQLT 384

There is also homology to SEQ ID 1028.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 409

A DNA sequence (GBSx0444) was identified in *S.agalactiae* <SEQ ID 1325> which encodes the amino acid sequence <SEQ ID 1326>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3195(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1327> which encodes the amino acid sequence <SEQ ID 1328>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2735(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 38/55 (69%), Positives = 48/55 (87%)

```

Query: 8 DEFKEAIDKGYISGNTVAIVRKNGKIFDYVLLHEEVREEVVTVERVLDVLRKLS 62
      DEFK+ AID GYI+G+TVAIVRK+G+IFDYVL HE+V+ EVVT E+V +VL +LS
Sbjct: 5 DEFKQAIDNGYIAGDTVAIVRKDGQIFDYVLPHEKVKNGEVVTKKVEEVLVELS 59

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 410

A DNA sequence (GBSx0445) was identified in *S.agalactiae* <SEQ ID 1329> which encodes the amino acid sequence <SEQ ID 1330>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1331> which encodes the amino acid sequence <SEQ ID 1332>. Analysis of this protein sequence reveals the following:

-515-

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 57/66 (86%), Positives = 63/66 (95%)

Query: 1 MSQEKLSKSLDQAKGGAKEGFGKITGDKELEAKGFIEKTIAGKELADDAKDAVEGAVDA 60
 MS+EKLKSK++QA GG KEG GK+TGDKELEAKGF+EKTIAGKELADDAK+AVEGAVDA
Sbjct: 1 MSEELKSKIEQASGGLKEGAGKLTGDKELEAKGFVEKTIAGKELADDAKEAVEGAVDA 60

15 Query: 61 VKEKLLK 66
 VKEKLLK
Sbjct: 61 VKEKLLK 66

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 411

A DNA sequence (GBSx0447) was identified in *S.galactiae* <SEQ ID 1333> which encodes the amino acid sequence <SEQ ID 1334>. This protein is predicted to be TnpA (orfB). Analysis of this protein
25 sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3961(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9667> which encodes amino acid sequence <SEQ ID 9668>
35 was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1335> which encodes the amino acid sequence <SEQ ID 1336>. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 152/160 (95%), Positives = 154/160 (96%)

50 Query: 1 MKNMALPKMATVKKTKALKTKTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 60
 MKNMALPKMATVK KALK+TKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL
Sbjct: 194 MKNMALPKMATVKKTKALKRTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 253

45 Query: 61 CAIIDLYSRKYIAWKLSHRMDAKLACDTLELALNKRKIEGTLFLHSDQGSQFKAREFRKI 120
 CAI+DLYSRK IAWKLSHRMDAKLACDTLELALNKRKIEGTLFLHSDQGSQFKARE RKI
55 Sbjct: 254 CAILDLYSRKCIWKLSHRMDAKLACDTLELALNKRKIEGTLFLHSDQGSQFKARELRKI 313

-516-

Query: 121 IDNNIMHSFSKPRYPYDNAVTEAFFKYLKHRQINQKNYQ 160
 IDN IMHSFSKP YPYDNAVTEAFFKYLKHRQINQK YQ
 Sbjct: 314 IDNNTIMHSFSKPGYPYDNAVTEAFFKYLKHRQINQKKYQ 353

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 412

A DNA sequence (GBSx0448) was identified in *S.agalactiae* <SEQ ID 1337> which encodes the amino acid sequence <SEQ ID 1338>. Analysis of this protein sequence reveals the following:

10 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1090(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 413

A DNA sequence (GBSx0449) was identified in *S.agalactiae* <SEQ ID 1339> which encodes the amino acid sequence <SEQ ID 1340>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.57 Transmembrane 17 - 33 (6 - 38)
 INTEGRAL Likelihood = -4.67 Transmembrane 147 - 163 (142 - 166)

30 ----- Final Results -----

bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 178/435 (40%), Positives = 281/435 (63%), Gaps = 1/435 (0%)

40 Query: 1 MKKLKIFPKMFIQIFSLGILILVHSLFFFIFPKTYLTKRVKIHIMADEISKNMNGKE 60
 MK+ +F K+FI FSI +L+I +H +F+FP TYL R+ I A I++++ GK+
 Sbjct: 1 MKRTGLFAKIFIYTFISIFSVLIVICLHLAIYFLPSTYLSHRQETIGQKATAIAQSLEGKD 60

45 Query: 61 LKYLDQTLLELYSKSSDIKVFIIKNNKNELQINDNINNVKSDSNLSLIEEREIKLHDGK 120
 + ++Q L+LYS++SDIK +K +++L++ D++ ++ + SL IEERE+K DG
 Sbjct: 61 RQSIQVLDLYSQTSDIKGTVKGEMTEDKLEVKDSLPLDTRQTTSLFIEEREVKTQDGG 120

Query: 121 KIHLDQFVSTADMQKDAKDLSLKFLPYSLISFLSIVISLIYAKSIKNNIQEITMVTDKM 180
 + LQF+++ D+QK+A+ +SL+FLPY+L SFL S++++ IYA++I I EI VT +M
 50 Sbjct: 121 TMILQFLASMDLQKEAQISLQFLPYTLASFLISLLVAYIYARTIVAPILEIKRVTRM 180

Query: 181 IKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLEKLYDFFKGAS 240
 + LD + RL++ S DEIG LK+QIN LY LL I DL KN+ IL+LEK+K +F +GAS
 Sbjct: 181 MDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNBAILQLEKMKVEFLRGAS 240

-517-

Query: 241 HELKTPLSLKKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQILSFYSIKDLNND 300
 HELKTPPL+SLKIL+ENM+ NIG+YKDRD Y+ + IVD L +V QILS S+++L +D
 Sbjct: 241 HELKTPPLSLKILLENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSLSSVQELRDD 300

Query: 301 EEYLNVGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVFSNLISNAVKYTN 360
 E +++ +++ Y++L ++++ I+ L Y+ + + ++ SNLISNA+K++
 Sbjct: 301 RETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLLISNLISNAIKHSV 360

Query: 361 RNGIINIKIANDWLLIENSYDKNKISKINKILDASFDLKLDNSNGLNIVKNILNKYNI 420
 G++ I L IENS + K+ + + K+ S G+GL +VK++L +
 Sbjct: 361 PGGLVIRIGEREGELFIENSCSSSEEQEKLAQSFSDNASRKVKGS-GMGLFVVKSLLEHEKL 419

Query: 421 KYEILHGENYFIFKI 435
 Y EN F I
 Sbjct: 420 AYRFEMEENSLTFFI 434

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1341> which encodes the amino acid sequence <SEQ ID 1342>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.83 Transmembrane 14 - 30 (6 - 35)
 INTEGRAL Likelihood = -2.44 Transmembrane 157 - 173 (156 - 174)

----- Final Results -----
 bacterial membrane --- Certainty=0.5734 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 123/455 (27%), Positives = 223/455 (48%), Gaps = 23/455 (5%)

Query: 3 LIKKTFVLVINGLIIVVTSILLVLVYFAMPIYTKVKDKCEFDQTSKQIKGKTIVTEIR 62
 L K F+ + V+V + L +YF P Y + + + + ++ ++GK I
 Sbjct: 6 LFAKIFIYTFISFVLVICHLLATYFLPSTYLSHRQETIGQKATAIAQSLEGKDRQSIE 65

Query: 63 DILTKKINKNDNIWYSLVSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKVM 122
 +L +I ++ ++ L++ D + D Q ++ +
 Sbjct: 66 QVLDLYSQTSDIKGTV---KGEMTEDKLEVKDSLPLDTRQTTSLF-----IEE 111

Query: 123 QKVTLRDGGKMTLLGQSSLQPVTDASKVLLDLVPSLLIFSVTVGSIVAYLYSRTSSRRIL 182
 ++V +DG M L +S+ +A ++ L P L+ S + +VAY+Y+RT IL
 Sbjct: 112 REVKTQDGGTMILQFLASMDLQKEAQISLQFLPYTLLASFLISLLVAYIYARTIVAPIL 171

Query: 183 SMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRIYASLSTSIKSLQKEYEKASDSERE 242
 + + ++M++L+ + + KDEI L IN LY L T I L ++ E E+
 Sbjct: 172 EIKRVTRRMDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKM 231

Query: 243 KSEFLRMTSHELKTPITSVIGMIDGMLNVGDFADRDKYLKRCRDVLEGQAQLVQSILSL 302
 K EFLR SHELKTP+ S+ +I+ M N+G + DRD+YL +++ V ILSL
 Sbjct: 232 KVEFLRGASHELKTPPLASLKILLENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSL 291

Query: 303 SKIETLASQNQLFSLKSSLEEMEVLVLSELKHLKVTINLEBQFVKANKVYLLKAIKN 362
 S ++ L ++E L + + + + +L++ + L++ +L Q N + + N
 Sbjct: 292 SSVQEL-RDDRETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLLISN 350

Query: 363 IIDNAFHYTKSGGQVMIQLKDNQLVIKNEAETLLTQQQMKQLFQPFYRDPYSRNRKGGT 422
 +I NA ++ GG V I ++ +L I+N + ++ ++L Q F + +RK G+
 Sbjct: 351 LISNAIKHSVPGGLVRIGEREGELFIENSC---SSEEQEKLAQSF---SDNASRKVKGS 403

Query: 423 GLGLFITHQILDQHHLAYRFVVLQDQRMVFTIDFP 457
 G+GLF+ +L+ LAYRF +++ + F IDFP
 Sbjct: 404 GMGLFVVKSLLEHEKLAYRF-EMEENSLTFFIDFP 437

-518-

An alignment of the GAS and GBS proteins is shown below:

Identities = 108/454 (23%), Positives = 220/454 (47%), Gaps = 22/454 (4%)

```

5  Query: 4  LKIFPKMFIQIFSIILGILILVHSLFFFIKPKTYLETRKVKIHIHMADEISKNMNGKELKY 63
      +++ K F+ I ++ +++ + + +F P Y + + ++ D+ SK + GK +
      Sbjct: 1  VRLIKKTFVLVINGLIIVVVTSLILVLVLFAMPIYYTKVKDKCEVKCFDQTSKQIKGKTUTE 60

10 Query: 64  LDQTLELYSKSSDIKVFIIKNNK-----NELQINDNINNVKSDSN--SLII 109
      + L +I + ++N+ +E + + N+N+ D++ ++ +
      Sbjct: 61  IRDILTKKINKDNIWYSLVSDNQQLYPSLQLLDGVSEKDSQNVNIVTTFDNSYSNVKV 120

15 Query: 110  EEREIKLHDGKKIHLQFVSTADMQKDAKDLKFLFYSLSISFLSIVISLIYAKSIKNN 169
      +++ L DGKK+ L S+ DA + L P L S +++ +Y+++
      Sbjct: 121  MSQKVTLRDGGKMTLLGQSSLPQVTDASKVLLDLYPSSLIFSVTVGSIVAYLYSRTSSRR 180

20 Query: 170  IQEITMVTDKMIKLDKETRLKISSNDEIGQLKQINDLYCALLNTINDLEFNKEILKLE 229
      I ++ KM+ L+ I DEI L IN LY +L +I L+ + ++ E
      Sbjct: 181  ILSMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSE 240

25 Query: 230  KLYDFFKGASHELKTPSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQIL 289
      + K +F + SHELKTP++S+ +++ M YN+G + DRD Y+ +C ++++ + V IL
      Sbjct: 241  REKSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFAADRDKYLKCRDVLEGQAQLVQSIL 300

30 Query: 290  SFYSIKDL-NNDEEYLNVDLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVF 348
      S I+ L + ++E ++ +L+E +E + +L K + + L + + K L
      Sbjct: 301  SLSKIETLASQNELFSLKSSLEEMEVLVLSELKHLKVTINLEEQFVKANKVYLLKAI 360

35 Query: 349  SNLISNAVKYTNNGLIINIKIANDWLLIENSYDKNKISKINKILDASF-----DLKLDN 402
      N+I NA YT G + I++ ++ L+I+N + + K L F + D
      Sbjct: 361  KNIIDNAPHYTKSGGQVMIQLKDNQLVIKNEATLLTQQQMQLFQFPYRDPYSRNRKDG 420

      Query: 403  SNGGLNIVKNILNKYNIKE-ILHGENYFIFKI 435
      GLGL I IL++++ Y ++ + + +F I
      Sbjct: 421  GTGLGLFITHQILDQHHLAYRFVVLQQRWVFTI 454

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 414

40 A DNA sequence (GBSx0450) was identified in *S.agalactiae* <SEQ ID 1343> which encodes the amino acid sequence <SEQ ID 1344>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence
45  INTEGRAL    Likelihood = -0.80    Transmembrane    50 - 66 ( 50 - 66)

----- Final Results -----
      bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9665> which encodes amino acid sequence <SEQ ID 9666> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

55  >GP:AAD25108 GB:AF140356 VncR [Streptococcus pneumoniae]
      Identities = 131/218 (60%), Positives = 176/218 (80%), Gaps = 1/218 (0%)

      Query: 5  MKILTVEDDKLIREGISEYLSBFGYTVIQAQDGRREALSKFNS-DINLVILDIQIPFINGL 63

```


-519-

MKIL VED+++IREG+S+YL++ GY I+A DG+EAL +F+S ++ LV+LDIQ+P +NGL
 Sbjct: 1 MKILIVEDDEMIREGVSDYLTDCGYETIEAADGQEALEQFSSYEVALVLLDIQMPKLNGL 60
 Query: 64 EVLKEIRKKSNNLPILILTAFSDEEYKIDAFNLVDGYVEKPFSLPVLKARIDSLIKKNFG 123
 EVL EIRK S +P+L+LTAF DEEYK+ AF +L DGY+EKPFSL +LK R+D++ K+ +
 Sbjct: 61 EVLAEIRKTSQVPVLMMLTAFQDEEYKMSAFASLADGYLEKPFSLSLKVRVDAIFKRYD 120
 Query: 124 HLEKFEYKNLSVNFNSYTAKEINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDS 183
 F YK+ V+F SY+A + +++ +NAKELEIL L+ N+G+ LTR QIID VWK +
 Sbjct: 121 TGRIFSYKDTKVDSESYASLAGQEVPIINAKELEILDYLVKNEGRALTRSQIIDAVWKAT 180
 Query: 184 EEIPYDRVVDVYIKELRKKLQDCITTIRNVGYKLERK 221
 +E+P+DRV+DVYIKELRKKL LDCI T+RVNGYKLERK
 Sbjct: 181 DEVPPDRVIDVYIKELRKKLQDCILTVRNVGYKLERK 218

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1345> which encodes the amino acid sequence <SEQ ID 1346>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 48 - 64 (48 - 64)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF72358 GB:AF192329 VanRB [Enterococcus faecalis]
 Identities = 88/215 (40%), Positives = 128/215 (58%), Gaps = 2/215 (0%)
 Query: 3 KILVVEDDDTISQVCEFLKANNYDPCVDFGQAALDKWQTTSYDLIILDIMLPSLSGLE 62
 +IL+VEDDD I + FL Y D DG A K+ +Y L+ILDIMLP ++G E
 Sbjct: 4 RILLVEDDDHICNTVRGFLAEAGYQVDACTDGNAYTKFYENTYQLVILDIMLPGMNGHE 63
 Query: 63 VLKTIKRTSDVPIIMLTALDDEYTLQVSNFNLISDYVTKPFSPILIKRIENVLRVSTPD 122
 +L+ R +D PI+M+TAL D+ Q+ +F+ DYVTKPF IL+KR+E +LR S
 Sbjct: 64 LLREFRAKNVTPILMMTALSDDENQIRAFDAEADDYVTKPFKMQILLKRVEALLRRSGAL 123
 Query: 123 EKR-QIGDLLVDETEHSVYWQTLVKLTKEYDIIDYLAKRHQKIVTRDQLMDDIWGYS- 180
 K ++G L + + +V GT + LT+KE++I+ L + + +T + ++ IWGY
 Sbjct: 124 AKEIRVGRILTLLPEDFTVLCDTELPLTRKEFEILLVQNGKRTLTHEILSRIWGYDF 183
 Query: 181 ELDTRVLDNHIKNLRKKMTGIPLKTTITGMGYLLGE 215
 E D + HIKNLR K+ +KTI G+GY L E
 Sbjct: 184 EGDGTVHTHIKNLRKLPENIIKTIRGVGYRLEE 218

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/214 (37%), Positives = 126/214 (58%), Gaps = 4/214 (1%)
 Query: 6 KILTVEDDDKLIREGISEYLFSEFGYTVIQAQDGREALSKFNS-DINLVILDIQIPFINGLE 64
 KIL VEDD I + I E+L Y DG+ AL K+ + +L+ILDI +P ++GLE
 Sbjct: 3 KILVVEDDDTISQVCEFLKANNYDPCVDFGQAALDKWQTTSYDLIILDIMLPSLSGLE 62
 Query: 65 VLKKEIRKKSNNLPILILTAFSDEEYKIDAFNLVDGYVEKPFSLPVLKARIDSLIKKNFGH 124
 VLK IRK S++PI++LTA DE ++ +F +L+ YV KPFS +L RI+++++ +
 Sbjct: 63 VLKTIKRTSDVPIIMLTALDDEYTLQVSNFNLISDYVTKPFSPILIKRIENVLRVSTPD 122
 Query: 125 LEKFEYKNLSVNFNSYTAKEINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDSE 184
 EK + +L V+ ++ + + KE +I+ L +++TR Q++D +W SE
 Sbjct: 123 -EKRQIGDLLVDETEHSVYWQTLVKLTKEYDIIDYLAKRHQKIVTRDQLMDDIWGYSE 181
 Query: 185 EIPYDRVVDVYIKELRKKLQDCITTIRNVGYKL 218
 RV+D +IK LRKK+ + TI +GY L
 Sbjct: 182 --LDTRVLDNHIKNLRKKMTGIPLKTTITGMGYLL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 415

- 5 A DNA sequence (GBSx0451) was identified in *S.agalactiae* <SEQ ID 1347> which encodes the amino acid sequence <SEQ ID 1348>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

Possible site: 49

10 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.68	Transmembrane	423 - 439 (413 - 447)
INTEGRAL	Likelihood = -10.67	Transmembrane	16 - 32 (12 - 37)
INTEGRAL	Likelihood = -9.77	Transmembrane	303 - 319 (301 - 326)
INTEGRAL	Likelihood = -3.13	Transmembrane	343 - 359 (343 - 367)

15 ----- Final Results -----

bacterial membrane	---	Certainty=0.6074(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
Identities = 280/458 (61%), Positives = 363/458 (79%), Gaps = 3/458 (0%)

25 Query: 1 MIKNAYVTRKSLKSLIIILVILSMATLSIISLSIKDATDRASKETFANITNSFSMEIN 60
M+ NAFAYVTRK KS++I L+IL MA+LS++ LSIK AT +AS+ETF NITNSFSM+IN
Sbjct: 1 MLHNAYVTRKFFKSIVIFLIILMASLSLVGLSIKGATAKASQETFKNITNSFSMQIN 60

30 Query: 61 RQVNPQTPRGGGNVKGEDIKKISQTSNIDSIVKRINSVADLVHDHIIETQDTLANQSPER 120
R+VN GTPRG GN+KGEDIKKI++ +I+SYVKRIN++ DL +D+IET +T N + +R
Sbjct: 61 RRVNQQTPRGAGNIKGEDIKKITEKAIKESYVVRINAIGDLTGVDLIETPETKKNLTADR 120

35 Query: 121 AKNFKRTVMLTGVNDSAKETKFVSEAYKLVEGKHLENKDKNKLMLHKDLAKNNLKVGD 180
AK F ++M+TGVNDS+KE KFVS +YKLVEG+HL N DK+KIL+HKDLA K+ KVGDK
Sbjct: 121 AKRFGSSSLMITGVNDSKEDKFVSGSYKLVEGEHLTNDKDKILLHKDLAAKHGKVGDK 180

40 Query: 181 IKIKSNLFPDADNEKVANETVEVEIKGLFDGHNSGGVSAAQELYENTLITDVHSAKVYGN 240
+K+ SN++DADNEK A ETVEV IKGLFDGHN V+ +QELYENT ITD+H+AAK+YG
Sbjct: 181 VKLDSNIYDADNEKGAKETVEVTIKGLFDGHNSKSAVTYSQELYENTAITDIHTAAKLYGY 240

45 Query: 241 TEDTAVYQDATFFVKGDKNLDSVIKDL-GKLDINWREYNLIKSSSNYPALQQSISGIYSI 299
TEDTA+Y DATFFV DKNLD V+K+L G INW+ Y L+KSSSNYPAL+QSISG+Y +
Sbjct: 241 TEDTAIYGDATFFVTADKNLDDVMKELNGISGINWKSYYTLVKSSSNYPALQESISGMYKM 300

50 Query: 300 SNKLFVGSLLIFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIEMVFISIPAL 359
+N LF GSL F+ ++++LLL LW+NAR+KE+ +LLS+G+ + I GQFI E + I+IPAL
Sbjct: 301 ANLLFWGSLSFVLLALLLSLWINARRKEVGILLSIGLKQASILGQFITESILIAIPAL 360

55 Query: 360 LGSYFLAQYTADKLGNILNKVTGDIKQIARQSASSQLGGGAEBGFNKTLSGLDINV- 418
+ +YFLA YTA +GN +L VT +AKQ ++ + +S LGGGAE +GF+KTL S LDI++
Sbjct: 361 VSAYFLANYTARAIGNTVLANVTSGVAKQASKAAQASNLGGGAEBVDFGFKTLSSLDISIQ 420

Query: 419 LPKFIIYVVFMSFVLLVSLILSSIYTLRKNPKELLID 456
FII V+ + V+LV + L+S LRK PKELL+D
Sbjct: 421 TSDFIIFVLALVLVVLV-MALASSNLLRKQPKELLID 457

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1349> which encodes the amino acid sequence <SEQ ID 1350>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

-521-

INTEGRAL	Likelihood = -12.90	Transmembrane	19 - 35 (16 - 43)
INTEGRAL	Likelihood = -7.27	Transmembrane	371 - 387 (359 - 392)
INTEGRAL	Likelihood = -7.01	Transmembrane	335 - 351 (326 - 357)
INTEGRAL	Likelihood = -6.21	Transmembrane	282 - 298 (276 - 308)

----- Final Results -----

bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]

Identities = 176/408 (43%), Positives = 250/408 (61%), Gaps = 16/408 (3%)

Query: 1 MENWKFALSSIIGHKMRSLTLMGLIIGVAAVVIIMGLGNAMKNSVTSTFSSKQKDIQLY 60
+EN + ALSS+ HKMRSLTLMGLIIGV +V++++ +G + + + S ++LY
Sbjct: 4 LENIRMALSSVLAHKMRSLTLMGLIIGVGSVIVVAVGQGGEQMLKQSIGPGNTVELY 63

Query: 61 FQEKGBE--EDLYAGLHTHENHNEVKEPWELEQIVKIDIDGIDSYFFINSATSTISYEKKKV 118
+ EE + A + +++K +K I+GI + S + Y +++
Sbjct: 64 YMPDEELASNPNAEAESTFTENDIKG-----LKGIEGKQVVASTSES MKARYHEEET 117

Query: 119 DNASIIGVSKDYFNKINYDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYSALNK 178
D A++ G++ Y N+ + I +GRT TDND+ +R+ ++ +A +LF K S L +
Sbjct: 118 D-ATVNGINDGYMNVNLSKIESGRTFTDNDFLAGNRVGIISQMAKELEFDK--TSPLGE 173

Query: 179 VVSLSDKDYLVIGVYKTDQTPVSPDGLSGGAVMANTQVASEFGTKEIGSIYIHVNDIQNS 238
VV ++ + +IGV K +SFD LS V N + S FGT + ++ + V +
Sbjct: 174 VVWINGQPVEIIGVLKKVTGLLSPD-LSEMYVFPN-MMKSSFQTSDFSNVSLQVESADDI 231

Query: 239 MNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISLLVGGIGV 298
+ G +AA L N +H + Y V + +I I +IMTT+IGSIA ISLLVGGIGV
Sbjct: 232 KSAGKEAAQ-LVNDNHGTEDSYQVMNMBEIAAGIGKVTAIMTTIIGSIAISLLVGGIGV 290

Query: 299 MNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGIGLIGLLAQLSVGALGN 358
MNIMLVSVTERTREIG+RK+LGATR +IL+QFLIESVVLTVLGGIGLIGLLAQLSVGALGN
Sbjct: 291 MNIMLVSVTERTREIGIRKSLGATRGQILTQFLIESVVLTVLGGIGLIGLIGVGGIGIG-YGGAALVS 349

Query: 359 AMTLKGACISLDVALIAVLFSASIGVFFGMLPANKASKLDPREALRYE 406
A+ + IS V VLFS IGV FGMLPANKA+KLDPIREALRYE
Sbjct: 350 AIAGWPSLISWQVCGGVLFMSLIGVIFGMLPANKAAKLDPIREALRYE 397

An alignment of the GAS and GBS proteins is shown below:

Identities = 56/247 (22%), Positives = 101/247 (40%), Gaps = 42/247 (17%)

Query: 147 YKLVEGKHLENKDKNKI-----LMHKDLAKKNNLK-----VGDKIKIKSNLFDA 190
Y +V G+ L + D + ++ DL K N K + DK + ++
Sbjct: 136 YDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYSALNKVVSLSDKDYLVIGVYKT 195

Query: 191 DNEKVANETVEVEIKGLFDGHNSSGVSAAQELYENTLITDVHSAKVYGNTEDTAVYQDA 250
D V+ FDG + G V A NT + A +G E ++Y
Sbjct: 196 DQTPVS-----FDGLSGGAVMA-----NTQV-----ASEFGTKEIGSIYIHRV 232

Query: 251 TFFVKGDKNLDSVIKDL--GKLDINWREYNLIKSSSNYPALQQSISGIYSISNKLFGVSL 308
++ NL + D+ I +Y + +S + S + ++ + SL
Sbjct: 233 ND-IQNSMNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISL 291

Query: 309 IFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIEMVFISIPALLGSYFLAQY 368
+ G+ V ++ + + R +EI + +LG ++L+I QF+IE V +++ L LAQ
Sbjct: 292 LVGGIGVMNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGIGLIGLLAQL 351

Query: 369 TADKLG 375
+ LGN
Sbjct: 352 SVGALGN 358

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 416

A DNA sequence (GBSx0452) was identified in *S.agalactiae* <SEQ ID 1351> which encodes the amino acid sequence <SEQ ID 1352>. This protein is predicted to be Vexp2 (b0879). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3194 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
Identities = 142/207 (68%), Positives = 169/207 (81%)

Query: 1 MDILEIKNVNYSYANSKEKVLGSGVNQKFELGKFYAIVGKSGTGKSTLLSLLAGLDKVQTG 60
M +L++++V Y Y N+ E VL +N FE GKFY+I+G+SG GKSTLLSLLAGLD G
Sbjct: 1 MTLQLQDVTYRYKNTAEAVLYQINYNFEPGKFPYSIIGESGAGKSTLLSLLAGLDSFVEG 60

Query: 61 KILFKNEDIEKKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSEILFELGLDK 120
ILF+ EDI KKGYS HR ++ISLVFQNYNLIDYLSPIENIRLVNKSVDSEILFELGLD+
Sbjct: 61 SILFQGEDIRKKGYSYHRMHISLVFQNYNLIDYLSPIENIRLVNKKASKNTLLELGLDE 120

Query: 121 KQIKRNVMLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKELAQDRN 180
QIKRNV++LSGGQQQRVAIAR+LVS+AP+ILADEPTGNLD TAG+I+ +LK LAQ
Sbjct: 121 SQIKRNVQLSGGQQQRVAIARSLVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTG 180

Query: 181 KCVIVVTHSKEVADSADIILELSGKKL 207
KCVIVVTHSKEVA ++DI LEL KKL
Sbjct: 181 KCVIVVTHSKEVAQASDITLLELKDKKL 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1353> which encodes the amino acid sequence <SEQ ID 1354>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2717 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 83/230 (36%), Positives = 135/230 (58%), Gaps = 13/230 (5%)

Query: 1 MDILEIKNVNYSYANSKEKVLGSGVNQKFEL--GKFYAIVGKSGTGKSTLLSLLAGLDKVQ 58
M +E+K V+ SY + V + FE+ G+ I+G SG GKST+L++L G+D V
Sbjct: 5 MAFIELKQVSKSYQIGETTTFANHEVSFEINKGELVVILGASGAGKSTVLNLLGGMDTV 64

Query: 59 TGKILFKNEDIE---KKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSE--- 111
G+++ +DI K + +R+N I VFQ YNL+ L+ EN+ L + V ++
Sbjct: 65 AGQVIIDGKDIAHYTSKALTQYRRNAIGFVFQFYNLVPNLTAKEVELAVEIVADALDPV 124

Query: 112 -ILFELGLDKKQIKRNVMLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIIN 170
IL E+GL + + +LSGG+QQRV+IARAL + ++L DEPTG LD T +I+
Sbjct: 125 TILKEVGLSHR-LDHFFAQLSGGQQQRVSIARALAKNPKLLLCDEPTGALDYQTGKQILT 183

-523-

Query: 171 ILKELAQDRNKCIVVTHSKEVADSADIILELSGKKLKK--VNKMNLEVE 218
 +L+++AQ + V++VTH+ +A AD ++ + ++ K +NK +E
 Sbjct: 184 LLQDMAQTGKTTVVIVTHNAIAPIADRVIFMHDAQVTKTVINKEPASIE 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 417

A DNA sequence (GBSx0453) was identified in *S.agalactiae* <SEQ ID 1355> which encodes the amino acid sequence <SEQ ID 1356>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.35 Transmembrane 17 - 33 (17 - 34)

----- Final Results -----
 bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 418

A DNA sequence (GBSx0454) was identified in *S.agalactiae* <SEQ ID 1357> which encodes the amino acid sequence <SEQ ID 1358>. This protein is predicted to be Vexp1. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 (285 - 312)
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 (385 - 417)
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 (14 - 38)
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 (333 - 357)

----- Final Results -----
 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47592 GB:AF140784 Vexp1 [Streptococcus pneumoniae]
 Identities = 165/425 (38%), Positives = 271/425 (62%), Gaps = 4/425 (0%)

Query: 2 IKNAIAYITRKKNRTLIIIFAILTIVLSCLYSLTIMKSSNEIEKALYESNSSISITK-K 60
 I+ + AY++RK+ R+ I+F IL ++L+ + +CLT+MKS+ +E LY+S N+S SI K +
 Sbjct: 4 IQRSWAYVSRKRLRSFILFLILLVLLAGISACITLMKSNKTVESNLYKSLNTSFSIKKIE 63

Query: 61 DGKYFNINQFKNIEKIKEVEEKIFQYDGLAKLKDVKVVSGEQSINREDLSDEFKNVVSLE 120
 +G+ F ++ ++ KIK +E + + +AKLKD + V+GEQS+ R+DLS N+VSL
 Sbjct: 64 NGQTFKLSDLASVSKIKGLENVSPLETVAKLKDKAEVGTGEQSVERDDLAAADNNLVSLT 123

Query: 121 ATSNTRKNNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKLLGDEIDLELLDTEKS 180
 A ++ +++ F+S F+ KEG++++ D IL+HEE AK+N L L D+I L+ +E S
 Sbjct: 124 ALEDSSKDVTFSTSAFNLKEGRHLQKGDSSKILIEELAKNGLSLHDKIGLDAGQSE-S 182

-524-

Query: 181 GGIKSHKFKIIGIFSGKKQETVTGLSSDFSENMFVVDYSTSQEILNKSENNRANKILMY 240
 GK ++ +F+IIGIFSGKKQE +TGLSSDFSEN VF DY +SQ +L SE A + Y
 Sbjct: 183 GKGQTVFELIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTLLENSEAQVSAARF--Y 240

Query: 241 SGSLESTELALNKLKDFKIDKSKYSIKKDNKAFEESESVSGIKHIKIMTYSIMLGGIV 300
 + + + + +++++ ++ Y ++K+NKAFF+ +SV+ + + I Y +++ G
 Sbjct: 241 VENPKEMDGLMKQVENLALENQGYQVEKENKAFEQIKDSVATFQTFLTIFLYGMLIAGAG 300

Query: 301 VLSSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFISIPSIISLFLGNLLKVI 360
 L L+L LWLRRER+YE+GI L++G K I QF E++ +S+ +++ + GN + +
 Sbjct: 301 ALILVLSLWLRERVYEVGILLALGKGSIFLQFCLEVVLVSLGALLPAFVAGNAITTYL 360

Query: 361 VEGFINSENSMIFGSLINKSSFMLNITTLAESYLILISIIVLSVVMASLILFKPKKEI 420
 ++ + S + +L SS +I + AESY+ L+ + LSV + + K PKEI
 Sbjct: 361 LQTLASGDQASLQDTLAKASSLSTSLSPAESYVFLVLLSCLVALCFLFLFRKSPKEI 420

Query: 421 LSKIS 425
 LS IS
 Sbjct: 421 LSSIS 425

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1359> which encodes the amino acid sequence <SEQ ID 1360>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.57	Transmembrane	23 - 39 (16 - 43)
INTEGRAL	Likelihood = -11.36	Transmembrane	371 - 387 (362 - 396)
INTEGRAL	Likelihood = -8.12	Transmembrane	331 - 347 (324 - 360)
INTEGRAL	Likelihood = -7.70	Transmembrane	280 - 296 (277 - 308)

----- Final Results -----
 bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB97962 GB:U96166 ATP-binding cassette transporter-like protein
 [Streptococcus cristatus]
 Identities = 222/311 (71%), Positives = 278/311 (89%)

Query: 16 MRSILTMLGIIIGIGAIIAIFSIIIEGNTENTKRQLIGGSNNTINIVFNKSSIDPKFPDK 75
 MRS+LTMGLGIIIGIGAIIAIFSIIIEGNTENTKRQLIGGSNNTI +V++KKS+IDP P+K
 Sbjct: 1 MRSMLTMLGIIIGIGAIIAIFSIIIEGNTENTKRQLIGGSNNTIKVVYDKKSAIDPSIPEK 60

Query: 76 SNAKKPDYLPFMAEEELSКИQVKGVKNALISYGIDDKVYHLGQKSSAKISAITKNVAEV 135
 S A+KP Y+PFM E+ LSKI+++ GVKNAL++YG D+K+Y+L QKSS+K+ A++++VA++
 Sbjct: 61 SQAQKPSYIPFMGEDVLSKIKEIPGVKNALMTYGADEKIYYLSQKSSSKVQAVSQSVADI 120

Query: 136 RRMTFIKGSDFSDKDFIDQKQVIYLEKSLEYSLFPKDDGLGKFVEVMGNPFRVIGVFESK 195
 ++ ++G F + F +Q+QV YLEKSLY++LFPK DG+GK+VEV GNPF+VIGVFES
 Sbjct: 121 KQQRLLLEGEGFDSEAFKNQEQVAYLEKSLYDTLFPKGDGIGKYVEVKGNPFKVGVPFEST 180

Query: 196 EQSGLTSGTEKIAIYIPLHQWYNINGVVDATPEITIQTYRADDLKPVAKRVSDMLNQITPK 255
 EQSGLTSG+EK+AYIPL QW+ I ++ +PE+T+QT++ADDLK VAK+VSD LNQ +P+
 Sbjct: 181 EQSGLTSGSEKVAYIPLQWHRIFDTINVSPEVTQTHKADDLKVKVSDYLNQQMPQ 240

Query: 256 SDYMFVGMNLKEFERQLDNLNKNFVLLAGIASISLIVGGIGVMNIMLVSVTERTREIGI 315
 SDYMFGV+NL+EFERQLDNLN+SNFVLLAGIASISL+VGGIGVMNIMLVSVTERTREIGI
 Sbjct: 241 SDYMFVGLNLQEFERQLDNLNQSNFVLLAGIASISLLVGGIGVMNIMLVSVTERTREIGI 300

Query: 316 KKALGARRKLI 326
 KKALGARRK++
 Sbjct: 301 KKALGARRKIL 311

An alignment of the GAS and GBS proteins is shown below:

-525-

Identities = 79/386 (20%), Positives = 170/386 (43%), Gaps = 38/386 (9%)

Query: 5 AIAYITRKNRTLIIFAILTIVLSCLYSLTIMKSSNE-IEKALYESSNSSISITKKGK 63
 A++ I K R+++ + I + + + +I++ + E ++ L SN++I+I
 5 Sbjct: 7 ALSSILSHKMRSILTMLGIIIGIGAITAIFSIIEGNTENTKRLIGGSNNTINIV----- 61

Query: 64 YFNINQFKNIEKIKEVEEKIFQYDGLAKLKDVKVSGEQSINREDLSDEFKNVVSLEATS 123
 FN K ++ K F AK D E+ +++ KN +
 10 Sbjct: 62 -FN-----KKSSIDPK-FPDKSNAPDYLFFMAEEELSKIQQVKGVKNALISYGID 111

Query: 124 NTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFKQNKLLGDEIDLELDTE----- 178
 + +L S KN+ E + + + +F+ ++ + I LE E
 Sbjct: 112 DKVYHLGQKSSAKISAITKNVAEVRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPK 171

Query: 179 -----KSGKIKSHKFKIIGIFSGKKQETTYTGLSSDFSENMFVVDYSTSQEILNKSENNRI 233
 K ++ + F++IG+F K+Q +GL+S +E + ++ I +
 15 Sbjct: 172 DDGLGKFVEVMGNPFRVIGVFESKEQ---SGLTSG-TEKIAYIPLHQWYNINGVVDATPE 227

Query: 234 ANKILMYSGSLESTELALNKLKDFKIDKSKYSIKDN-KAFEEESLESVSGIKHIK--IM 290
 + L+ ++ + + I KS Y N K FE L++++ ++ I
 20 Sbjct: 228 ITIQTYRADDLKPVAKRVSDMLNQTIPKSDYMFVGNLKEFERQLDNLNKS NFVLLAGIA 287

Query: 291 TYSIMLGGIVVLSLILWLRLRIYEIGIFLSIGTTKIQIIRQPIFELIFIS----IPSI 346
 + S+++GGI V++++L+ + ER EIGI ++G + I++QF+ E + ++ + +
 25 Sbjct: 288 SISLIVGGIGVMNIMLVS-VTRETRIGIKKALGARRKLILKQFLIEAVILTLLGGVIGV 346

Query: 347 ISSLFLGNLLKLVIVEGFINSSENSMI 372
 IS + G ++ + + +I S S++
 30 Sbjct: 347 ISGMVSGLIITRSLEYPIILSLFSVV 372

A related GBS gene <SEQ ID 8571> and protein <SEQ ID 8572> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 McG: Discrim Score: 5.59
 35 GvH: Signal Score (-7.5): -5.97
 Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 4 value: -11.52 threshold: 0.0
 40 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 (285 - 312)
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 (385 - 417)
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 (14 - 38)
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 (333 - 357)
 PERIPHERAL Likelihood = 4.51 315
 modified ALOM score: 2.80
 45 *** Reasoning Step: 3
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

38.7/67.3% over 421aa
 55 Streptococcus
 pneumoniae
 GP|5712667| Vexpl Insert characterized

ORF00815(304 - 1575 of 1875)
 60 GP|5712667|gb|AAD47592.1|AF140784_1|AF140784(4 - 425 of 425) Vexpl {Streptococcus
 pneumoniae}
 %Match = 25.0
 %Identity = 38.7 %Similarity = 67.2
 Matches = 164 Mismatches = 136 Conservative Sub.s = 121
 65

-526-

	48	78	108	138	168	198	228	258
	SIEH*VVFDNKTI*T*ELDFVSHSS*VI*DFPLNK*IRNSVTSYINGSIIBIVCQMKKP*WK*F*KH*L*AM*KY*SSG							
5	288	318	348	378	408	438	468	495
	CNSCGVKIERSN*EVIKNAIAYITRKQNRTLIIPAILTIVLSCLYSCLTIMKSSNEIEKALYESSNSSISITK-KDGKYF : : :: : ::: :: : : :: : : : : :: : MNPIQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKS NKTVESNLKYSLNTSFSIKKIENGQTTF 10 20 30 40 50 60							
10	525	555	585	615	645	675	705	735
	NINQFKNIEKIKEVEEKIFQYDGLAKLKDLYVSGEQBSINREDLSDEFKNVVSLEATSNTKRNLFFSGGVPSFKEGKNIE : : : : : : : : : :: : :: :: : : : : : :: ::: KLSDLASVSKIKGLENVSPLETIVAKLKDKAEAVTGEQSVERDDLAAADNNVLVSLTALEDSSKDVTTFTSSAFNLKEGRHLQ 80 90 100 110 120 130 140							
15	765	795	825	855	885	915	945	975
	ENDKNSILVHEEFAKONKLGDEIDLELDTEKSGKIKSHKFKIIGIFS GKQETYTGLSSDFS ENMVVDYSTSQEIL : :: : :: : : : : : : : :: : : : KGD SKKILI HBE LAKNGLSLHD KIG LDAG QSE-SGKGQTV EFEL IIGIFS GKQ EKFTGLSSDFS ENQVFTDYESSQTL 160 170 180 190 200 210 220							
20	1005	1035	1065	1095	1125	1155	1185	1215
	NKSENNRRIANKILMYSGSLESTELALNKLKDFKIDKSKYSIKKDNKAFEESLESVSGIKHIHKIMTYSIMLGGIVVLSLI : : : : : : : : : : : : : : : : : : GNSEA--QVSAARFYVENPEKMDGLMKQVENLALENQGYQVEKENA FEIQIKDSVATFQTFLTIFLYGM LIA GAGALILV 240 250 260 270 280 290 300							
25	1245	1275	1305	1335	1365	1395	1425	1455
	LILWLRERIYEIGIFLSIGTTTKIQIRQFIFELIFISIPSIISSLFLGNLLKVIVEGFINSSENSMIFGGSLINKSSFML :: :: :: :: : :: :: :: :: : : : : : : : : : LSLWLRERVYEVGTLLALGKGKSSI FLQFCLEVVLVSLGALLPAFVAGNAITTYLLQTL LASGDQASLQDTLAKASSLST 320 330 340 350 360 370 380							
30	1485	1515	1545	1575	1605	1635	1665	1695
	NITTLAESYLILISIIVLVSVMASSLILFKPKPILSKIS*EQIMDILEIKNVNYSYANSKEKVL SGVNQKFELGKFYAI : : : : : : : : : SILSFABESYVFLVLLSCLVALCFLFLFRKSPKEILLSIS 400 410 420							
35	1725	1755	1785	1815	1845	1875	1905	1935
	NIITLAESEYILISIVLVSVMASSLILFKPKPILSKIS*EQIMDILEIKNVNYSYANSKEKVL SGVNQKFELGKFYAI : : : : : : : : : SILSFABESYVFLVLLSCLVALCFLFLFRKSPKEILLSIS 440 450 460							

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 419

A DNA sequence (GBSx0455) was identified in *S.agalactiae* <SEQ ID 1361> which encodes the amino acid sequence <SEQ ID 1362>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 42
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL      Likelihood = -5.04      Transmembrane      19 - 35 ( 14 - 42)

----- Final Results -----
50      bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-527-

Example 420

A DNA sequence (GBSx0456) was identified in *S.agalactiae* <SEQ ID 1363> which encodes the amino acid sequence <SEQ ID 1364>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
5  >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 421

A DNA sequence (GBSx0457) was identified in *S.agalactiae* <SEQ ID 1365> which encodes the amino acid sequence <SEQ ID 1366>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
20 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA74029 GB:U30715 ORFB [Bacillus anthracis]
Identities = 33/76 (43%), Positives = 44/76 (57%), Gaps = 1/76 (1%)
30
Query: 11  IRRVSHACKAGDRFYEEINILNREFTATAHNQKWCTDVTYLQYGLGAKAYLSAIKDLYNG 70
      ++R      R  EN+LNR F A  N+KW TD+TYL +G      YL +I DLYN
Sbjct: 86  VKRKRRTWINGESRIVVENLLNRNFQANKPNEKWVTDITYLPFGT-EMLYLLSIMDLYNN 144

35
Query: 71  SIIAYEISHNNEIHL 86
      IIAYEIS+  ++ L+
Sbjct: 145  EIIAYEISNRQDVTLV 160

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 422

A DNA sequence (GBSx0458) was identified in *S.agalactiae* <SEQ ID 1367> which encodes the amino acid sequence <SEQ ID 1368>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
45 >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.69    Transmembrane    10 - 26 ( 10 - 26)

----- Final Results -----
      bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
50     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 423

A DNA sequence (GBSx0459) was identified in *S.agalactiae* <SEQ ID 1369> which encodes the amino acid sequence <SEQ ID 1370>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4170(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA56999 GB:U09558 ORFA, putative Helix-Turn-Helix motif from
      amino acid 21 through 42 and from amino acid 78 through
      99 [Lactobacillus johnsonii]
Identities = 28/116 (24%), Positives = 59/116 (50%), Gaps = 6/116 (5%)

Query: 3   YSTLAKEQGVQGYLDGKGLSLRDICKWYDISSRSVLQKWIKRYTSGEDLKATSRGYSRMKQ 62
      YST K + V YL+ + S++ + K Y+I +++++W+ + + L A S +++
Sbjct: 4   YSTELKIEIVSKYLNHEDSIKGLAKQYNIHW-TLIRRWVDK-AKCQGLAALSVKHTKTTY 61

Query: 63   GRQATFEERVEIVNYTIAHGKDYQAAIEKFGVSYQQIYSWVRKLEKNGSQGLVDRR 118
      + ++ +V Y + H          KF +S Q+Y+W +K + G GL+ ++
Sbjct: 62   SS----DFKLNVVRYLTHSIGVSKVAAKFNISDSQVYNWAKKFNEEGYAGLLPKQ 113
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 424

A DNA sequence (GBSx0460) was identified in *S.agalactiae* <SEQ ID 1371> which encodes the amino acid sequence <SEQ ID 1372>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -0.69    Transmembrane    2 - 18 ( 2 - 19)
```

```
----- Final Results -----
      bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 425

A DNA sequence (GBSx0461) was identified in *S.agalactiae* <SEQ ID 1373> which encodes the amino acid sequence <SEQ ID 1374>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:AAC79517 GB:U88974 ORF1 [Streptococcus thermophilus temperate
       bacteriophage 01205]
       Identities = 104/172 (60%), Positives = 127/172 (73%), Gaps = 11/172 (6%)

   Query: 10  QHQSYAALYLIAKTGMRFAECLGLTVNDIDYINKYLSINKTWDYHFNQRYLPTKNKSSIR 69
              ++ SYAALY+I+KTG+RFAECLGLTV+DI      LS+NKTWDY  N  ++PTK KSSIR
20   Sbjct: 186 EYASYAALYIISKTGIRFAECLGLTVDDIKRDTGMLS VNKTWDYKNNNGFMPPTKTKSSIR 245

   Query: 70  NIPIDNDTLFFLHEFTKNKNDRLFDKLSNNAVNKTIRKITGREVRVHSLRHTFASY---- 125
              IP+D++ + F+ +      + RL  LSNNAVNKT+RKI GREVRVHSLRHT+ASY
25   Sbjct: 246 EIPLDDEFINFDQLPPTDDGRLLPSLSNNAVNKTIRKITVGREVRVHSLRHTYASYLIAH 305

   Query: 126 ---LISISQVLHENLNITLEVYAHQLQEQRNDKLNQRNLGRIWGKIALN 174
              LIS+SQVL HENLNITLEVYAHQLQEQRND+      + ++W K  N
30   Sbjct: 306 DIDLISVSQVLGHENLNITLEVYAHQLQEQRNDE----KIKQMWTKCGQN 353

```

30 There is also homology to SEQ ID 578

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 426

35 A DNA sequence (GBSx0462) was identified in *S.agalactiae* <SEQ ID 1375> which encodes the amino acid sequence <SEQ ID 1376>. Analysis of this protein sequence reveals the following:

```

   Possible site: 22
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.3206(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1328.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 427

50 A DNA sequence (GBSx0463) was identified in *S.agalactiae* <SEQ ID 1377> which encodes the amino acid sequence <SEQ ID 1378>. Analysis of this protein sequence reveals the following:

possible site: 45

-530-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.6542(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB52541 GB:AJ131519 hypothetical protein [Lactobacillus
 bacteriophage phi adh]
 Identities = 24/55 (43%), Positives = 36/55 (64%)

 Query: 12 MDKELTPQEKANKKWAENNREHRTYLSKRSTARSFINKNATKEDLLELKQLIESK 66
 M K + KANKKW E N+ + Y++KRSTA+SFI AT+EDL +++ + +
 15 Sbjct: 1 MAKITEARAKANKKWDEKNKARKLYINKRSTAKSFILNLATEEDLANIEEYVAER 55

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 **Example 428**

A DNA sequence (GBSx0464) was identified in *S.agalactiae* <SEQ ID 1379> which encodes the amino acid sequence <SEQ ID 1380>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4417(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1332.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 **Example 429**

A DNA sequence (GBSx0465) was identified in *S.agalactiae* <SEQ ID 1381> which encodes the amino acid sequence <SEQ ID 1382>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

40

----- Final Results -----

 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-531-

Example 430

A DNA sequence (GBSx0466) was identified in *S.galactiae* <SEQ ID 1383> which encodes the amino acid sequence <SEQ ID 1384>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -4.30    Transmembrane  205 - 221 ( 202 - 223)
    INTEGRAL    Likelihood = -3.56    Transmembrane  296 - 312 ( 294 - 312)

----- Final Results -----
10      bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9663> which encodes amino acid sequence <SEQ ID 9664> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8573> and protein <SEQ ID 8574> were also identified. Analysis of this protein sequence reveals the following:

```

20  Lipop: Possible site: -1    Crend: 8
    McG: Discrim Score:      -8.80
    GvH: Signal Score (-7.5): -4.03
    Possible site: 47
    >>> Seems to have no N-terminal signal sequence
25  ALOM program  count: 2 value: -4.30 threshold: 0.0
    INTEGRAL    Likelihood = -4.30    Transmembrane  205 - 221 ( 202 - 223)
    INTEGRAL    Likelihood = -3.56    Transmembrane  296 - 312 ( 294 - 312)
    PERIPHERAL  Likelihood = 2.97      20
    modified ALOM score: 1.36
30
    *** Reasoning Step: 3

----- Final Results -----
35      bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 SEQ ID 8574 (GBS366) was expressed in *E.coli* as a GST-fusion product. The purified fusion protein (Figure 215, lane 5) was used to immunise mice. The resulting antiserum was used for FACS (Figure 281), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 431

45 A DNA sequence (GBSx0467) was identified in *S.galactiae* <SEQ ID 1385> which encodes the amino acid sequence <SEQ ID 1386>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```

Possible site: 31
    >>> Seems to have no N-terminal signal sequence
50
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1471(Affirmative) < succ>

```

-532-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 A related GBS nucleic acid sequence <SEQ ID 8575> which encodes amino acid sequence <SEQ ID 8576> was also identified. This has an RGD motif at residues 81-83.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
 [bacteriophage Dp-1]
 Identities = 99/140 (70%), Positives = 120/140 (85%)
 Query: 10 MVLINIEQAIAMMASRKGVITYSMIDYRNGPSSYDCSSSVYFALRSAGASDNGWAVNTEYEH 69
 M ++IE+ +AWM +RKG+V+YSMD+R+GP SYDCSSS+Y+ALRSAGAS GWAVNTEY H
 Sbjct: 1 MGVDIIEKGVAVWQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60
 15 Query: 70 DWLIKNGYVLIAENTNWNQQRGDFIWIWGRGASAGAFGHTGMFVDPDNIHCNYGYNISIT 129
 WLI+NGY LI+EN W+A+RGDIFIWG++GASAGA GHTGMF+D DNIIHCNY Y+ I+
 Sbjct: 61 AWLIRNGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIHCNYAYDGIS 120
 20 Query: 130 VNNHDEIWIWYNGQPYVYAYR 149
 VN+HDE W Y GQPY Y YR
 Sbjct: 121 VNDHDERWYIYAGQPYVYVYR 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1387> which encodes the amino acid sequence <SEQ ID 1388>. Analysis of this protein sequence reveals the following:

25 Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.06 Transmembrane 79 - 95 (77 - 95)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 56/91 (61%), Positives = 68/91 (74%)
 Query: 158 KVDNQSVSVSKFEKELDVNTPLSNSNMPYYEATISEDYYVESKPDVNSTDKELLVAGTRVR 217
 K+D F ++LD NT L NSN+PYYEAT+ DYYVESKP+ +S DKE + AGTRVR
 40 Sbjct: 354 KIDKPKSQSLTFNQKLDINTKLDNSNVPPYYEATLRDYYVESKPNASSADKEFIKAGTRVR 413
 Query: 218 VYEKVKWARIGAPQSNQWVEDAYLIDATDM 248
 VYEKV GW+RI A QS+QWVED YL +AT +
 45 Sbjct: 414 VYEKVGWSRINASQSDQWVEDKYLSNATQV 444

SEQ ID 8576 (GBS301) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 9; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 3; MW 55kDa).

50 The GBS301-GST fusion product was purified (Figure 205, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 300), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 432

A DNA sequence (GBSx0468) was identified in *S.agalactiae* <SEQ ID 1389> which encodes the amino acid sequence <SEQ ID 1390>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 53
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -6.53    Transmembrane    8 - 24 ( 3 - 25)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 433

A DNA sequence (GBSx0469) was identified in *S.agalactiae* <SEQ ID 1391> which encodes the amino acid sequence <SEQ ID 1392>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 34
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
25          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 434

A DNA sequence (GBSx0470) was identified in *S.agalactiae* <SEQ ID 1393> which encodes the amino acid sequence <SEQ ID 1394>. Analysis of this protein sequence reveals the following:

```

35      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.0120(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 435

A DNA sequence (GBSx0471) was identified in *S.agalactiae* <SEQ ID 1395> which encodes the amino acid sequence <SEQ ID 1396>. Analysis of this protein sequence reveals the following:

Possible site: 38
 5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9661> which encodes amino acid sequence <SEQ ID 9662> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 436

20 A DNA sequence (GBSx0472) was identified in *S.agalactiae* <SEQ ID 1397> which encodes the amino acid sequence <SEQ ID 1398>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.39 Transmembrane 349 - 365 (347 - 366)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF43531 GB:AF145054 ORF39 [Streptococcus thermophilus
 bacteriophage 7201]
 Identities = 212/666 (31%), Positives = 323/666 (47%), Gaps = 52/666 (7%)

35 Query: 10 WGNLTLLEILSAWNKP---NIASNTSTVNVQVFL-----KMSSYGYISIGETRPLKITVD 61
 W NN + W +I +NTS V +++ L + Y + E ++
 Sbjct: 5 WSNNDRGYRIRLWVDQVGQDIQNNTSQVRLRLSLLNTTTTFAQYSCSAFVEFNGQRLNWS 64

40 Query: 62 GRAETINVNPSINYGQRKLLFAKDYIVNHNSDGNKPLFNISAYYPIN--FSNYGEATANQ 119
 G + N +I L + V H DG+ +F + A++ + +S NQ
 Sbjct: 65 GSPSVLGMNQTIQ-----LIDQTTVRHADDGSG-VFGVHAHFNGSGGWSFGNLDIGNQ 117

45 Query: 120 SISLPKINRLSVSSAISGVLGNAVITITINRYSTSFTHNLKYDFKGSTGTIATGVGTSYLV 179
 I+L I R S G +GN V I+I+R TH L+Y ++ G IA VGTSY W
 Sbjct: 118 QITLTTPRGSSVRVSDGFIGNQVDISIDRKIGGATHTLRYAWENKQKGIADNVGTSYKW 177

Query: 180 TIPPTFANLLPNELTGTGNLIVETMDGSAKIGETKYTLSTIPNTATYKPKLSSITLSDT 239
 TIP FAN +PN +G G + V+T I TL+ ++ T KP + TL+DT
 50 Sbjct: 178 TIPEDFANDIPNSTSGRGTIYVDITYINGNFIQTQSTLTASV-ITNNLKPSFTGFTLTDT 236

Query: 240 NTLTSSIVSG-NNFVRIISKVKVDFGSAIGNNGSTITSYNAEIVGKSNSIIGNSVFDKL 298
 N + IV G +FV I+S VKV F A +G+TI Y AEIVG +NSI NG V ++
 Sbjct: 237 NPTSQRIVPGQTHFVSIMSLVKVVFNGAQAKSGATIVGYAETVGANNSSISSNGGVLREV 296

-535-

Query: 299 DFFGSA--TIRATVTDSEPTSEPVDTKINVIDYFLPIVTSKVVRSQQNPDLQVLPFV 356
 T+R V DSRG+ S+ V+TK+ + YF P + +V RS + DIL + F
 Sbjct: 297 SVNQDTEMTLRGRVQDSRGISWDVETKLTFLFYFSPAL-RFEVKRSDKKLDILTIRFA 355

Query: 357 KIAPIIIVGGIQQNQLKMSVSVAPYNTGIYAVDSGAATNTWSTISQMSGAPLNLGGTYDKS 416
 KIAP+ V GIQ+N +K++ S A + VD+G A WS+IS+ + + LG +Y
 Sbjct: 356 KIAPLSVNGIQRNVNMLTFSTAKVGWDFVVDNGQAGGVWSSISEFNASDAKLGNSTPAD 415

Query: 417 KSWLVKISVSDNLSATPIIQPVASEFVLVTKAPSGVAFGKIWEHGIIDAKGDVYVDGTI 476
 S++V + D S T V ++ V++T GV GK E G +D GD I
 Sbjct: 416 TSYVVIGKLEDEFTS-TSFQATVPTDEVIMTYDRQGVGIGKYRERGALDVNGD-----I 468

Query: 477 YCGDKAIQQKPLALNNGGSFRHDDTDLNSLQDTGFYCVFRGANRPAGAGPGYVTVVRHET 536
 Y + IQQ L NNG ++ N+++D G Y +F A P + + H +
 Sbjct: 469 YANNSPIQQYQLTNNNGSPKMTNNA--NTIEDPGQYYLFSAA--PGNPSGQWGHLEFHHSS 524

Query: 537 -----ANYAYQQFYDRTNKTI-----FTRLLENGVWVGWSEYVKKD--SLQTTGWITIG 583
 A Q F+ + ++R++++ W W E+ + D +L TGW G
 Sbjct: 525 YGKGSYKKEAIQIFWSNDGRLFSRHRWSRIIDD--WEPWKEFARNDNTNLINTGWQPAG 582

Query: 584 -NGFKYKRKGDDIDLMYNFASNGLQRWSVGNMPSGLI--POELMFAITGWTLPDKSIHL 640
 +G YKR GD + + +NF G + + ++P + PQ MF +TGW++ +K ++
 Sbjct: 583 VDGsfYKRVGDVLTIKFNFTGTG-GDFLLASVPPEIFKAPQSYMFVVVGWSVWANKQYNV 641

Query: 641 QINASG 646
 Q+N G
 Sbjct: 642 QVNEGG 647

30 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1398 (GBS365) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 2; MW 102kDa).

GBS365-GST was purified as shown in Figure 216, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 437

A DNA sequence (GBSx0473) was identified in *S.agalactiae* <SEQ ID 1399> which encodes the amino acid sequence <SEQ ID 1400>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3481(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34413 GB:AF158600 putative minor structural protein
 [Streptococcus thermophilus bacteriophage Sfil1]
 Identities = 504/998 (50%), Positives = 675/998 (67%), Gaps = 56/998 (5%)

Query: 1 MLTIHGPDLPVLFELDNDKQCALNYFNHKWYRKQKTGSSVLEFSVYKDLLGDSPLSHKY 60
 +LTIH +L+ V ++DN+KQ LN+FN KW R ++G+SV EFSV+KK + DS + Y
 Sbjct: 2 LLTIHDNNLQKVAYIDNEKQSTLNFFNDKWTRSLSEGSTSVFEFSVFKKSIKSDSKVEISY 61

Query: 61 HVLNDQAFVSVFVHKGKVQLLNIMKIDEDEKQIDCYCENLNLELLNEYCNAYKATKAMSFE 120

-536-

LN++AFVSF HKGK L N+MKI+EDE+ I CYCENL+LELL EY AYKA+K M+F+
 Sbjct: 62 KYLNERAFVSFKHKGKSYLNFVMKIEEDEQIIRCYCENLSLELLLEYRGAYKASKPMTFK 121

5 Query: 121 EYLVQFDILSWGALTIVGTNEVKDKKLTLEWTSQETKLARLLSIANNPDAEIEFETKLNFN 180
 EY + + + LT+G NEV D+K TLEW QET LARL+S+A NFAEIEF+T+L N
 Sbjct: 122 EYFDDWGMGQFAKLTILGVNEVSDQKRTLEWEGQETTLARLISLARNPDAEIEFDTRLKPN 181

10 Query: 181 HTFKQLIINIYKEYEKGKSYGVDRDKTDVILRYQKNISGIRKTVDKRQIYNNAIRPYGKK- 239
 + ++N+YK Y+ GK+ GV R ++DVIL+Y KNI+GI+++VDK QIYN I PYG+K
 Sbjct: 182 SQLDEFLNVYKAYD-GKNQGVGRRRSVDILKYGKNINGIKRSVDKTIYNMITYGRKS 240

15 Query: 240 -TVRGERVISNPVTRKVTKTGVSNT---YLGDLKYGHYTIKKANVQAIINYAVQYNIL 295
 T + + IS+PVT + V S R Y GGD L Y GHT+ + VQ I N VQ N+L
 Sbjct: 241 DTKKETKRISDPVTIQNPVVPSARVEKRYAGDLYAGHTLSASLVQTI FNL CVQRNLL 300

20 Query: 296 PSGIITQLYLESFWGDSTVGKRDNNWAGMSGGAQTRPSGVKVTGMAEPANEGGTYMHYA 355
 PSG+L+QLYLESFWG S V +RDNNW+GM+GGAQTRPSGV VTTG RPA+EGGTYMHYA
 Sbjct: 301 PSGVISQLYLESFWGSSNVARDDNNWSGMTGGAQTRPSGVVVTGSPRASEGGTYMHYA 360

25 Query: 356 SVDDFLKDYTYLLAKQG-----IYNVVGKKNIAADYTKGLFRAGGAKYDYAAAGYQSYTNL 410
 SVDDF+KDYTYLLA Q +Y V GK+NI +YTKGLFR GGA DYAAAGY Y L
 Sbjct: 361 SVDDFMKDYTYLLADQTSGGKRMVGVKGKQNIIEEYTKGLFRIGGALYDYAAAGYNHYIYL 420

30 Query: 411 MTNIRNGINKVTGNILNTIDKLWQTPVKPITAVNVARRATKTIQA-----INEATKLKG 464
 M +IRNGIN+ GNIL+ +D LW+ P IT N ++ T+T++A +NE LKG
 Sbjct: 421 MRDIRNGINRSNGNILDKLDLWRQPDNQITQPN--KQVTRTVKADRVIAVLNEMQGLKG 478

35 Query: 465 RRGISGQCICALSGWYAKLDGAWIDSSIGGIRGRIGGMAAALIGTDYNWGAYGWKVDKS 524
 RR+G+GQCICAL+ WY+ KL G + + + G G IG GMAAA IGTDY W +GW V +
 Sbjct: 479 RRVGNQCICALAAWYSMKLGGPGLGAGVTGKSGVIGAGMAAAKIGTDYAWDRFGWSVVRP 538

40 Query: 525 PNAGNLKAGGIYNVRANRGAPFYTTGWGHTGIIKSVSKTRVTIVLEQN FVGRMYVVENSYD 584
 + LK G I N++A T+ WGH II S + + VTVLEQN+ GR YVV+NSY
 Sbjct: 539 TSVDQLKPGAIAIANIKAYNSY-LGTSVWGHVSIIISNNGSTVTVLEQNYAGRQYVVQNSYP 597

45 Query: 585 INSFASGLQTVCYPREIAQGMVNGATTQQVSGGTQISYEEVVQEAQTESYEEBQIIYID 644
 +++ ++T+CYP E+ +G +V G T + ++ E+ + E + ID
 Sbjct: 598 ASAYLGAVERTLCTYPELKEGKTVEGRTEVSTPNVEVQKVEIPPIDVEVTTESTAALTID 657

50 Query: 645 NSIYKEWKDENGKVEYYLKNGLYAPLSRDRYPSVLTGNETRDNWIRKDMEVETDSQEV L 704
 + +EW++ENG+VE+YL+NG LYAP+S++ YPS+LTG E DNWIRKDME++TDS++VL
 Sbjct: 658 SKRKQEWNRNENGQVEFYLENGSLYAPISKELYPSILTGKENDNWKDMEIDTDSDEVL 717

55 Query: 705 MSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPLILTARVVEQEISITNPS 764
 +ST L++L+ YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV EQ+IS TNP
 Sbjct: 718 ISTALRNLRKFCYPAITYEVDGFLLDIGTQTKIQDTGFSPLMLLEARVSEQQISFTNPV 777

60 Query: 765 SNKTKFSNFVEKESQLASDLISDMLRLYDESIPYEIKLATSNGVAFKNGTGESVLTPLSQ 824
 NKT F+NF +++++ L+S M +L +E+IPYE+KL+T NG FKN TG+SVL +L+
 Sbjct: 778 ENKTVFANFQTLQNKVSDSLSRMTKLAEBAIPYELKLSTDNGTTFKNSTGQSVLKATLE 837

65 Query: 825 KNGKDYEAIFYKNGDSLIDIGPSLIVKASDFNHVNLITVEAYLNEELVASTQISFTDTE 884
 KNG+ Y+ ++F+KNGDS+I G L+VK +DF + L +TVEAYL++ELVAS +I+FTD
 Sbjct: 838 KNGEVYQPIFFKNGDSIIIGTGNQLVVKPTDFENTLQVTVEAYLDELVASAEITFTDVS 897

70 Query: 885 DEADGKDGAPGPQGPVGNLQCPKGDQGIQGPAGADGKATYTHIAYALDENGSTGFSVS 944
 DG QGPKGD G+ L S G+
 Sbjct: 898 DGK-----QGPKGDDGVS-----PINLIIESSNGYQFK 925

75 Query: 945 DNVGKTYI--GMVDDNIIDSNDPK-KYKWNLIKADG 979
 +N+ T +Y D+ ID + + Y W+ + ADG
 Sbjct: 926 NNIINTTFTAKLYQDNKEIDKDGTRYAYLWSKV-NADG 962

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1401> which encodes the amino acid
 65 sequence <SEQ ID 1402>. Analysis of this protein sequence reveals the following:

Possible site: 37

-537-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.56 Transmembrane 325 - 341 (323 - 343)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2423(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 23/55 (41%), Positives = 27/55 (48%)

Query: 886 GADGKDGAPGPGPPGVNGLQGPKGDQGIQGPAGADGKATYTHIAYALDENGSTG 940
 G GKDGPAG G PG G +G +G+ G QGP G G+ T G G

15 Sbjct: 181 GEAGKDGAPGKDGAPGEKGEKGRGETGAQGFVGPQGEKGETGAQGPAGPQGEAG 235
 Identities = 48/151 (31%), Positives = 58/151 (37%), Gaps = 19/151 (12%)

Query: 852 KASDFNHVLNITVEAYLNE--ELVASTQISFTDTEGDGKDGAPGPGPPGVNGLQGPK 909
 K DF L E L E +L+ + I + G G G GPQG G G QGPK

20 Sbjct: 82 KEEDFQKELKDFTEKRLKEILDICKSGIK---GDRGETGPAGPAGPQGTGERGAQGP 138

Query: 910 GD---QGIQGPAGADGKATYTHIAYALDENGSTGFS---VSDNVGKTYIGMYVDDNIID 962
 GD QGIQG AG G+ E G G + GK D

Sbjct: 139 GDRGEQGIQKAGEKGERGEKGDKGETGERGEKGEAGIQGPQGEAGK-----DGAPGK 191

25 Query: 963 SNDPKKYKWNLIKADGARGIQGPAGADGKT 993

P + +G GA+G GP G G+T

Sbjct: 192 DGAPGEKGEKGRGETGAQGPVGPQGEKGET 222

Identities = 25/50 (50%), Positives = 29/50 (58%), Gaps = 9/50 (18%)

30 Query: 884 EDGADGKDGAPGPGPPGVNGL-----QGPKGQGIQGPAGADGKA 924

+DGA GKDGPAG +G G G QG KG+ G QGPAG G+A

Sbjct: 185 KDGAPGKDGAPGEKGEKGRGETGAQGPVGPQGEKGETGAQGPAGPQGEA 234

35 SEQ ID 1400 was expressed in four different forms. SDS-PAGE analysis of total cell extract is shown in
 Figure 122 (GBS105dN - lane 5 & 7; MW 102kDa), Figure 122 (GBS105dC - lane 8-10; MW 81kDa),
 Figure 179 (GBS105d - lane 8; MW 102kDa) and in Figure 181 (GBS105C - lane 2; MW 56kDa).
 GBS105dN-His was purified as shown in Figure 232 (lanes 9 & 10). GBS105dC-His was purified as shown
 in Figure 233 (lanes 3 & 4).

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 438

A DNA sequence (GBSx0474) was identified in *S.agalactiae* <SEQ ID 1403> which encodes the amino
 acid sequence <SEQ ID 1404>. This protein is predicted to be a minor structural protein. Analysis of this
 protein sequence reveals the following:

45 Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2502(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34412 GB:AF158600 putative minor structural protein

55 [Streptococcus thermophilus bacteriophage Sfil1]

Identities = 163/433 (37%), Positives = 244/433 (55%), Gaps = 21/433 (4%)

-538-

Query: 80 LSSKKPKMLMFSHIPGRYYLAVQVGD LNFKEIKMNGFGEIT--FIVADAYAHSTSYRRIK 137
 L +KK L P RYLA+ G+++ K I + + E T F+V D AHST+Y+R+
 Sbjct: 93 LHTPKKAVKLFLPTEPERYXALVKGESLKGIS-DWYDEATIEFLVPDGVVAHSTTYKRVT 151

Query: 138 DYTQDGNKMTFKIKNNGTAPAFPIFRIKHLGNGYIGITNETGAFVGSPEEEDGTIVHR 197
 DY + KM F I N G+ A+PI +K ENGY G+ ++ AF G+ EE DG I+ +
 Sbjct: 152 DYQEKDGKMFISIDNEGSTDAYPIITLKANAENGYGLVSDKFAFEAGNIEADGKIISK 211

Query: 198 NETLFDY-SKAIAQAL-EGAPNVAKLNYMPPTFDSELKRMRLDNILSGSGGEYVAIGAR 255
 E L+D+ I QA +GA NV N + + + N+ G IG +
 Sbjct: 212 AEVLYDFRDDRIPQAFKAGAKNVGITINVTGDLHGT----LEIQNVWGRPH-----IGLK 261

Query: 256 GTTPGYGE-HVGTRTFIINPDSNGEY-TLNEHLWNKQIFIATAQDQKGFLLKLCVTGENDE 313
 + + T I PDS+G LNE++WW+QIF A + Q GFLKL V+ +
 Sbjct: 262 NPNANINQLQTASLTLDIPPDSGNGALNEYIWWRQIFWAGSISQYGLKLTVSDADGN 321

Query: 314 FLYGIETYKRKNGFETEYNFFALDDGVGWRFYKQFEFQA-DRNYHNPFSMNRSAVEIF 372
 FLYG+ET+KR G E+EYN A D G G+RF KQ+ F A + HNPF+ R + +I
 Sbjct: 322 FLYGVETFKRSLGLESEYNALASDGYG-GFRFLKQWSFLATEYEDHNPFPNEPRGWS-DIK 379

Query: 373 REEDKFRIYFNGAHHHTVPSLKGKKSRIHLAMGTCSDDSSKYINYNLFEKVNFEKMGVS 432
 RE+DK Y+ G ++ T+P +KGKKS KIHL + S ++ + F+++ + K +
 Sbjct: 380 REDDKVTFYWWGTNTFTIPEIKGKKSAKIHLTISNI-PSKSFVTHAYFDQLLYIKTNNA 438

Query: 433 HYNINIVNKYQPGDEVIINFENDTVSTKDIDSIQDVLGSKMISIPPGESSELVHLSSWVA 492
 + +I N+Y G +IIN E+DT++ ++ ++ ++V GS IPPGES++ V S W
 Sbjct: 439 FFEDIPNRYIQSSNLIINSEDDTLTLNLLNLDEIVDGSLLWVPVIPPGESQIEVVQSPWAK 498

Query: 493 ALPDISIDFEERY 505
 P ++I+FEER+
 Sbjct: 499 KKPSVTIEFEERW 511

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 439

A DNA sequence (GBSx0475) was identified in *S.agalactiae* <SEQ ID 1405> which encodes the amino acid sequence <SEQ ID 1406>. This protein is predicted to be PblA. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.11	Transmembrane	427 - 443 (424 - 445)
INTEGRAL	Likelihood = -4.99	Transmembrane	449 - 465 (448 - 469)
INTEGRAL	Likelihood = -2.71	Transmembrane	41 - 57 (38 - 57)
INTEGRAL	Likelihood = -0.37	Transmembrane	361 - 377 (361 - 377)
INTEGRAL	Likelihood = -0.22	Transmembrane	324 - 340 (324 - 340)

----- Final Results -----

bacterial membrane	---	Certainty=0.3845(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG18638 GB:AY007505 PblA [Streptococcus mitis]
 Identities = 233/401 (58%), Positives = 296/401 (73%), Gaps = 17/401 (4%)

Query: 1 MAINLGQAYVQIMPSAKGISGSIKTLDPASSAGSSAGSLGKLGILGVSIAAAKIG 60
 MAT + QAYVQ++PSA+GI+G I L+PEAS+AG SAG LG L+G++ VIAAA IG

-539-

Sbjct: 1 MATEIAQAYVQLIPSARGITGKIQSILNPEASAAGQSAGQSLGSSLVGVMTKVIAAAGIG 60

Query: 61 EMVTKAIISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
KA.S++ISEGAALQQSLGG+ETLTK +A+ VK YA+EAYKTTGLSANAYME+VTGF

5 Sbjct: 61 ----KAFSAAISEGAALQQSLGGIETLTKGSADKVKGYANEAYKTTGLSANAYMENVTGF 116

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFQKQNYTMLDNLKL 180
SASLLQSLGGDT KAA+ ANMAMIDM+DN+NKMGTSMESIQ AYQGFQKQNYTMLDNLKL

10 Sbjct: 117 SASLLQSLGGDTNKAETANMAMIDMSDNANKMGTSMESIQMAYQGFQKQNYTMLDNLKL 176

Query: 181 GYGGTQEEMKRLLSDAQKLTKKKYDISNLSDVYEAHAIQKGIGITGTTAKEAATFTGS 240
GYGGT++EM+RL+DA+KLTK KYDI+NLSDVY AHAIQ + ITGTAKEAA+TF+GS

Sbjct: 177 GYGGTKQEMQRLLDAAEKLTKGVKYDINNLSDVYSAHAIQENLDITGTTAKEAASTFSGS 236

15 Query: 241 FEAMKAASKNLLGKMGALGEDIKPSLKALPDTTSNFVLNNFIPMLTNVFKGFGSVISLTFS 300
FE+MKAA++N+LGK+ALGE+I PSL AL TTS F+ +NF+PM+ NVF G G V++ S

Sbjct: 237 FESMKAAQNVLGKLGALGENILPSLHALLKTTSTFLFDNFLPMIGNVFSGLGLVLTGEGIS 296

Query: 301 ELIPKIVGFMQTSQPSLMQSGISFIISFVNGFLTAYPAFLTAVAGKIFTDFVSFVMQSIPG 360
++ ++ G S + +S + G + F + G + ++ +I G

20 Sbjct: 297 QIASQLFG-----DAFGSAVFDQLSRITGIFETF--FDMIFGSLSKQDNIDILNTI-G 346

Query: 361 LLQAGATLVNLIDGILANLPQIATS---AVSVISSFISML 398
+ AT ++N+ D I I ++ V ++ F+ L

25 Sbjct: 347 FSEEAATQIVNIADNIRVTTFENIGSAIGDVVGIVGDFVGD 387
Identities = 112/386 (29%), Positives = 172/386 (44%), Gaps = 18/386 (4%)

Query: 235 TTFGTSGFEAMKAASKNLLGKMA-LGEDIKPSLKA---LFDTTSNFVLNNFIPMLTNVFKG 290
TT+ E++KA ++ +L E IK + L T V+ FI N++

30 Sbjct: 580 TTWNAVVESLKAMWNAVVTFFSDLWESIKEAATAWTLITTAVMMVQPFIDGFMNIWNN 639

Query: 291 FGSVISLTFSELIPKIVGFMQTSQPSLMQSGISFIISFVNGFLTAYPAFLTAVAGKIFTDF 350
++ + + G + S+ I II V G A L++ + +

35 Sbjct: 640 ISEGLTQVWEGIKLIFEAWFEI-KSIPLGAILIIDLVTGNFGQLGADLSLIWEGIKNG 698

Query: 351 VSFVMQSIFGLLQAGATLVNLIDGILANLPQIATS AVSVISSFISMLQANYPAILKKGF 410
+S + + I +++ G+ N + ++ I + SM + I

Sbjct: 699 ISLIWEGIKTYFSGVVDVIVGYATGVFENFSNVLSTIWEFIKTAASMA---WEWIKSTVS 755

40 Query: 411 EILSYLVQGIARLPDIVITVTKL---IAILAGAIASNLPKVLALGVQLLITFVKGILSV 467
+++ L+QG + V + L I A A S L K L LG + V G +

Sbjct: 756 NLITGLIQGAQNLWNNFVFSFLSGLWENIKSTASAAWSGL-KSLVLG--FINGLVSGAQT 812

Query: 468 IGKINETANNIGK---LINAISIDLLSAGRAIMRGFLRGLEDVWGDIONFVGDIAWGI 524
+ + +++ K + N IK+I+L AG+AI+ GFL GL+ W + NFVG IA WI

45 Sbjct: 813 WNNMKQAVSDLVTKVTNIFNGIKINLWEAGKAILNGFLGGLKSAWEGVTNMFVGGIANWI 872

Query: 525 KDHKGPISYDRRLIPAGNAIMQGLHQGLVDKFKPKVKNLVNGMAEEIQSSFCNPQLAFDM 584
+DHKGPI YDR+LLIPAGNAIM L GL D FK VK V GM+ EI F L +

50 Sbjct: 873 RDHKGPIEYDRKLLIPAGNAIMGSLDNLKDGFKDVKKTVGGMSEISDVFSGDNLDLNS 932

Query: 585 DTNVNNGFE-RIGTLNKNLSSQVTST 609
+V E R+ + L Q + T

55 Sbjct: 933 TASVTKNLEARLAMPQAQLEVQESKT 958

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1407> which encodes the amino acid sequence <SEQ ID 1408>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

60	INTEGRAL	Likelihood = -2.76	Transmembrane	458 - 474 (458 - 474)
	INTEGRAL	Likelihood = -2.60	Transmembrane	483 - 499 (482 - 499)
	INTEGRAL	Likelihood = -2.02	Transmembrane	429 - 445 (429 - 445)
	INTEGRAL	Likelihood = -1.28	Transmembrane	397 - 413 (397 - 413)
	INTEGRAL	Likelihood = -0.53	Transmembrane	739 - 755 (738 - 755)
65	INTEGRAL	Likelihood = -0.27	Transmembrane	356 - 372 (356 - 372)

-540-

----- Final Results -----

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AAB18717 GB:U38906 ORF42 [Bacteriophage rlt]
 Identities = 261/579 (45%), Positives = 359/579 (61%), Gaps = 63/579 (10%)

10 Query: 184 MKRLLSDAEKLPAAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTFSGSLAAMK 243
 M+RLL+DA+KL G+K+D+SN++D+ +AIH +Q M I G A+EA TTFSGS +MK
 Sbjct: 1 MQRLLTDAQKLT---GQKYDISNFSBITQAIHAIQTEMDITGTTAKEASTTFSGSFDSMK 57

15 Query: 244 SSFTNVMAGLSLGDDIRPALRGLAETTSNLFEGNFIPMVANIFKGLPSAIGTFFIGAAAPI 303
 ++ +NV+ LSLG D++ L L TTS FLF NFIPMV NIFK LP AI TF+ AA
 Sbjct: 58 AAMSNVLGNLSLGRDLQGLNALVSTTSTFLFKNFIPMVGNIFKALPGAISTFVSAAGKE 117

20 Query: 304 ITSQ-----FQGLMSSLG-ISIDLSPIT 325
 ++SQ F L+SS+G IS + +
 Sbjct: 118 LSSQLGNGIGSGFSDFETAKFSSILSPLQGSFQITIVSGLKPVFDSLLSSIGPISTQIMGVF 177

25 Query: 326 AKFAQIGQNLQ----PVFNGLKTAFSQLPSFFTSIGSAVAPVIDTIISGLARLDFSGFEA 381
 +K Q+ N+ PV + L AF QLPS F +I AV P+IDTI SG++RLDFSG +A
 Sbjct: 178 SKLPQLFSNVISAVIPVISTLSVAFGQLPSLFEATSVAVQPMIDTISSGSRDLDFSGIQA 237

30 Query: 382 LISAILPALQAGFSNFAIVGPAISGVVDSFVGMWNAAQPLISILSDALMPVFQILGSFL 441
 +ISA++PA+ G + I+GP+I +V+SFV MWN+ QPL ++++ ALMP FQ+LG+F+
 Sbjct: 238 IISALVPAITTGITTMGIIGPSIDTLVNSFVKMWNISIQPLATVIAGALMPAFQVLGAFI 297

35 Query: 442 GGVVKGALMGVSEAFDAVKVAIQLVPTPIIDLVLVQGLNFVQFVLSVIAEWIGVAIGMFGNL 501
 GGV+KGA++ +S FD ++V + +TPII ++ PVL+ +A+W+G AIG F N
 Sbjct: 298 GGVVLKGAMLALSATFTIRVVVGFLTPIIAAVLAKFQEFAPVLATVAQWVGTAIGFFANF 357

40 Query: 502 GTAGQGLSAFIKSAWNIQFAISTAGTIISTVIDYIKLAFSGAGSAVGVLKNIFSLAWMA 561
 G AG L I SAW I++ IS+ + I +I+ K F+G GSA G L+++ S AW
 Sbjct: 358 GAAGTSLKGLITSANGIKSIISVVSGIGGIINTAKAIFTGLGSAGGALRSMISGAWSG 417

45 Query: 562 MGDAINVAKGISSVINGIKSAFSSFS-----SLVSSVGSVAVNGVIDSISSTIRG--- 611
 + I+ G IS INGIKS FSS S++S V S + G+I SSTI G
 Sbjct: 418 IRSIISVGGISGTINGIKSFFSSLGSGNGLRSVMGVMGSGITGIISGASSTISGIID 477

50 Query: 612 -----LANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKGPISYDRVL 663
 L NID++GAG A++GF+ GLKS W A K FV GIA+WI +HKGPISYDR +
 Sbjct: 478 GIKNIFNSLKNIDLAGAGRAVIDGFGVGLKSTWEAGKKFVGGIADWIKDHKGPISYDRKI 537

55 Query: 664 LKPAGKAIMGGLNTSLIDGFKEVKSNNVSGMADDLASTMT 702
 L PAG+AIMGG N SL++ FK V+ NVSG+A + S +T
 Sbjct: 538 LIPAGQAIMGGFNDSLMENFKAVQKNVSGIAKQIQSAIT 576

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 272/701 (38%), Positives = 371/701 (52%), Gaps = 91/701 (12%)

55 Query: 1 MATNLGQAYVQIMPSAKGISGSISKTLDPASSAGSSAGSLGGLIGILGSVIAAAKIG 60
 MAT LQAYVQIMPSA+GISG+ISK LDPEA SAG SAGSL+GG L+ ++G IAAA IG
 Sbjct: 1 MATELGQAYVQIMPSARGISGAISKQLDPEARSAGLSAGSLIGGNLVKMIGGAIAAGIG 60

60 Query: 61 EMVTKAISSISSEGAALQQLGGVETLFPKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
 +M ISS++S GA LQQS CG++TL+K VK +A EAYK G+SAN Y E
 Sbjct: 61 KM----ISSALSAGADLQQSFGGIDTLYKGAETAVKGFAYKAYKA-GISANTYAEQAVSM 115

65 Query: 121 SASLLQSLGGDTAKAAKVNAMIDMADNSNKMGTSMESIQYAYQGFQKQNYTMLDNLKL 180
 ASL QSLGGD AAK ANMA++DMADNS KMGT + SIQ AYQGFQKQNYTMLDNL+L
 Sbjct: 116 GASLKQSLGGDAVAAAKAANMAIMDMADNSAKMGTDITSIQMAYQGFQKQNYTMLDNLRL 175

Query: 181 GYGGTQEEMKRLLSDAQKL---TGKKYDISNLSDVVEAIHAIQKIGITGTTAKEAATTF 237
 GYGGT+EEMKRLLSDA+KL GKK+D+SN +DV EAIH +Q +GI G A+EA TTF

-541-

Sbjct: 176 GYGGTKEEMKRLLSDAEKLPAA MGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTF 235

Query: 238 TGSFEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISL 297
+GS AMK++ N++ ++LG+DI+P+L+ L +TTSNF+ NFIPM+ N+FKG S I

5 Sbjct: 236 SGLAAMKSSFTNV MAGLSLGD DIRPALRGLAETTSNFLFCNFIPMVANIFKGLPSAIGT 295

Query: 298 TFSELIPIKIV---GFMQTSGLPSLMQSGISFIISFV-----NGFLTAY---PAFLTV 342
P I G M + G S+ S I+ + + NG TA+ P+F T

10 Sbjct: 296 FIGAAAPIITSQFQGLMSSLGISIDLSPITAKFAQIGQNLQPVFNGLKTAFSQLPSFFTS 355

Query: 343 AGKIPTDFVSFVMQSIPGL---LQAGATLVNLIDGILANLPQIATSAVS-VISSFISM 397
G + ++ + L +A + +L + +N I A+S V+ SF+ M

Sbjct: 356 IGSAPVVIDTIIISGLARLDFSGFEALISAILPALQAGFSNFAAIVGPAISGVVDSFVGM 415

15 Query: 398 LQANYPAI-----LKKGFELSYLVQGI-----IARLPDIVIT 430
A P I L F+IL + G+ + + D+++

Sbjct: 416 WNAAQPLISILSDALMPVFQILGSFLGGVVKGALMGVSFAFDVAVKVAIQLVTPIIDLLVQ 475

Query: 431 ----VGKLIAILAGAIASNLKPKVLALGV--QLLITFVKGILSVIGKINETANNIGEKLIN 484
V +++++A I + LG Q L F+K + I TA I +I+

20 Sbjct: 476 GLNFVQPVLSVIAEWIGVAIGMFGNLGTAGQGLSAFIKSAWTNIQTATAGTIIISTVID 535

Query: 485 AIKSI-----DILLSAGRAIMRGFLRGLEDVWGDIONFVGDI 521
IK D ++ + I+ + G++ + + V +

25 Sbjct: 536 YIKLAFSGAGSAVGVLKNIFSLAWMAMGDAINVAKGISSVINGIKSAFSSFSLSVSSVG 595

Query: 522 GWIKDHKGPISYDRLLI----PAGNAIMQGLHQGLVDKFKPKVKNLVNGMAEEIQSSFG 576
+ IS R L AG AIM G GL + VK+ V+G+A I G

30 Sbjct: 596 SAVNGVIDSISSTIRGLANIDISGAGAIMNGFLNGLKSAWGA VKSFVSGIANWIAEHKG 655

Query: 577 NPQLAFDMDTNVNNNGFERIGTLNKNLSSQVTSTDNYSGNA 617
+++D G +G LN +L + SG A

Sbjct: 656 --PISYDRVLLKPAGKAIMGGLNTSLIDGFKEVKSNVSGMA 694

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 440

A DNA sequence (GBSx0477) was identified in *S. agalactiae* <SEQ ID 1409> which encodes the amino acid sequence <SEQ ID 1410>. Analysis of this protein sequence reveals the following:

40 Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2565 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:AAG18637 GB:AY007505 unknown [Streptococcus mitis]
Identities = 64/119 (53%), Positives = 87/119 (72%), Gaps = 2/119 (1%)

Query: 1 MLKMD DALVCDLAET YHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
M++ DEDAL+CDLAETY I+DY+QLP +VAVF+ GLR++SRI ++ ++V FE +LA

55 Sbjct: 1 MIQTDEDALICDLAET YGIFDYRQLPADQVAVFAFGLRDDSRIKLAMTNSKVPFETFLLA 60

Query: 61 GMFDRLGMLIWMKTTDGGQKGNRPPEMVSTMF--DNQKQDSEVVSGSGKDFEETRNNIL 117
G+ DRL L+W KTTDGGQKG N+P MV+ + K+S+ + F SG+DFEE R IL

Sbjct: 61 GVLDRLSALVWFKTTDGGQKGINPLMVTEELTGKTKAKESKEMIFDSGEDFEEYRQKIL 119

60 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1411> which encodes the amino acid sequence <SEQ ID 1412>. Analysis of this protein sequence reveals the following:

-542-

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
5 bacterial cytoplasm --- Certainty=0.2905(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 60/123 (48%), Positives = 82/123 (65%), Gaps = 2/123 (1%)

Query: 1 MLKMDLEDALVCDLAETVHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
 M+ D+DAL CDLAETY IYDY+QLP +VAVF++GLR SRI +SG + + +LA
Sbjct: 1 MIAKDDDALTCDLAETVGIYDYRQLPAYQVAVFAVGLRSNSRIKMALSGETEALDTVLLA 60
15
Query: 61 GMFDRIGMLIWMKTTDGQKGKRNPEMV--STMFDNQKQDSEVVVSFGSGKDFEETRNNILG 118
 G++D +L W KT DGQ G+N+P+ V + QK ++V+SF SG+DFE R +LG
Sbjct: 61 GIYDNTNLLFWSKTKDGQSGQNKPKSVVREAISGSKSQKANDVISFVSGEDFENARKQLLG 120
20
Query: 119 FGG 121
 G
Sbjct: 121 GDG 123

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
25 vaccines or diagnostics.

Example 441

A DNA sequence (GBSx0478) was identified in *S.galactiae* <SEQ ID 1413> which encodes the amino acid sequence <SEQ ID 1414>. Analysis of this protein sequence reveals the following:

Possible site: 36
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2280(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG18636 GB:AY007505 unknown [Streptococcus mitis]
40 Identities = 40/80 (50%), Positives = 62/80 (77%), Gaps = 1/80 (1%)

Query: 3 TSSGFYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLGQVESLKNHLRASDGTVS 62
 TS+GF ++I + RL+NYEL+EA+++++NP LPKV++L+LG++ E LKNH+R +DG V
Sbjct: 24 TSTGFFFEITKERLENYELLEAISEVDTNPAVLKPKVKMLGNKSEDLKHNHVRTADGIVP 83
45
Query: 63 TEALMEEVKEIFES-GQLKK 81
 + + E+ EIF S QLKK
Sbjct: 84 LDKMGAEISEIFSSQNQLKK 103

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1415> which encodes the amino acid
50 sequence <SEQ ID 1416>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
55 bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-543-

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 42/75 (56%), Positives = 60/75 (80%)

5  Query: 2  KTSSGFYKIEESRLKNYELVEALADLESNPLSLPKVLRLLIGDQVESLKNHLRASDGTV 61
      KT+SGFEY+I + RLKN+ELVEA+A+ E++P ++ K++ LLLGD +SLK H+R ++G V
      Sbjct: 7  KTISGFYEIIPKRLKNFELVEAIAEEETDPTAVVKIVNLLIGDAKSLKEHVRDARGIV 66

      Query: 62 STEALMEEVKEIFES 76
      EA+ E+KEIFES
10  Sbjct: 67 DVEAIGVEIKEIFES 81

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 442

15 A DNA sequence (GBSx0479) was identified in *S.agalactiae* <SEQ ID 1417> which encodes the amino acid sequence <SEQ ID 1418>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence

20  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3461(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAG18635 GB:AY007505 unknown [Streptococcus mitis]
Identities = 114/183 (62%), Positives = 142/183 (77%)

30  Query: 2  VANSSNVTTAKPKIGGAIYTAPLGTLPKDTASELNEAFKSLGYISEDGLSNEDKRESEE 61
      +A +NVTAKPKIGGA+Y+APLGT LP D ++L++AF++LGYIS+DG++N + ESE
      Sbjct: 1  MATEANVTAKPKIGGAVYSAPLGTALPTDATTKLDQAFREALGYISDDGMTNSNSPESEN 60

      Query: 62 IQAWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYKDNVTGDLKTGITVKSNSKPLE 121
      I+AWGG VV S QK K D F Y LIEALN+ VLKE+YG DNV+GDL +GIT+K+NSK L
35  Sbjct: 61 IKAWGGVVVSSVQKEKTDTFKYMLIEALNLHVLKEVYGPDNVSGDLSSGITIKANSKELP 120

      Query: 122 EHCLVIEMILKNNNTVKRIVIPKGVSEVGEIKYVDNEAAGYETTLOAFPDAEGNTHYEYI 181
      HCLVIE +LK +KRIVIP GKV+ + EI Y D GY TT+ AFP+A +THYEYI
40  Sbjct: 121 HHCLVIETVLKGGVLRKIVIPSGKVTAIDEITYNDGSVLGYGTTVTAFPNAADDTHYEYI 180

      Query: 182 KGA 184
      KGA
45  Sbjct: 181 KGA 183

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1419> which encodes the amino acid sequence <SEQ ID 1420>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

50  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2379(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 119/182 (65%), Positives = 142/182 (77%)

```

-544-

Query: 4 NSSNVTAKPKIGGAIYTAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEETQ 63
 ++ NVT+AKPK GGAIY+APLGTETLPKD SELN FK+LGY+SEDG+ NED R SE I+
 Sbjet: 6 DTKNVTSAKPKTGGAIYSAPLGTETLPKDAKSELNTKFKNLGYVSEDGVVNEDTRSSSENIK 65

Query: 64 AWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYKDNVTGDLKTGITVKSNSKPLEEH 123
 AWGGD+V + Q K DKFTY LIE+LN+EVLKE+YG NVTGDL GI +KSNSK LE H
 Sbjet: 66 AWGGDIVGAVQTEKEDKFTYKLIESLNVEVLKEVYGAVNVTGDLGGGIQIKSNSKELEAH 125

Query: 124 CLVIEMILKNNTVVRIVIPKGVSEVGEIKYVDNEAAGYETTLQAFPDAGNTHYIYIKG 183
 +V++MI+ +KRIV+P KV EVGEIKYVD E GYETTL+ FPD +G+TH EYI
 Sbjet: 126 VIVVDMIMNGGILKRIVLPNAKVDEVGEIKYVDGEVVGYETTLKCFDPKDGDTREYIVK 185

Query: 184 AG 185
 G
 Sbjet: 186 PG 187

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 443

A DNA sequence (GBSx0480) was identified in *S.agalactiae* <SEQ ID 1421> which encodes the amino acid sequence <SEQ ID 1422>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2214(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18710 GB:U38906 ORF35 [Bacteriophage rlt]
 Identities = 52/78 (66%), Positives = 66/78 (83%)

Query: 1 MSKFKFKLNKAGVAELMKSSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
 M+K FKLN++GVA +MKS EMQ +L KA+A+++RCG GY QD+HVGKNRANAMV A+T
 Sbjet: 1 MAKNLFKLNRSVGASMMKSPMQAILKEKASAVKQRCGPGYGQDMHVGKNRANAMVFAET 60

Query: 61 IKAKKDNSKNNTLLKAVR 78
 +AK+DN KNNT+LKAVR
 Sbjet: 61 YQAKRDNMKNNTILKAVR 78

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1423> which encodes the amino acid sequence <SEQ ID 1424>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 75/78 (96%), Positives = 76/78 (97%)

Query: 1 MSKFKFKLNKAGVAELMKSSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
 MSKFKFKLN+AGVAELMKSSSEMQQVLTTKATAIRERCGDGY QDIHVGKNRANAMVS KT
 Sbjet: 1 MSKFKFKLNAGVAELMKSSSEMQQVLTTKATAIRERCGDGYVQDIHVGKNRANAMVSTKT 60

-545-

Query: 61 IKAKKDNSKNNLLKAVR 78
 IKAKKDNSKNNLLKAVR
 Sbjct: 61 IKAKKDNSKNNLLKAVR 78

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 444

A DNA sequence (GBSx0481) was identified in *S.agalactiae* <SEQ ID 1425> which encodes the amino acid sequence <SEQ ID 1426>. Analysis of this protein sequence reveals the following:

10 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAB18709 GB:U38906 ORF34 [Bacteriophage rlt]
 Identities = 41/59 (69%), Positives = 45/59 (75%)

Query: 1 MTGKKVEYILAIPKGDKHDWEDKEVCFDCKKRTVGLALEGIEELIPLEWNNKVMVERY 59
 +TGKK Y LAIPK D HDWE+K+V FF K WRT G LEGIE LIPL+WNKKV VE Y
 Sbjct: 56 LTGKKAIYTLAIPKKDTHDWENKKVRFFGKTWRTFGEPLLEGIEGLIPLDWNKKVTVEHY 114

25

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1427> which encodes the amino acid sequence <SEQ ID 1428>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

An alignment of the GAS and GBS proteins is shown below:

Identities = 51/60 (85%), Positives = 57/60 (95%)

40 Query: 1 MTGKKVEYILAIPKGDKHDWEDKEVCFDCKKRTVGLALEGIEELIPLEWNNKVMVERYE 60
 +TGKKVEY+LAIPKGD+HDWE+KEV FF KKWRTVG+ LEGIEELIPL+WNKKVMVERYE
 Sbjct: 50 LTGKKVEYVLAIPKGDHEDWENKEVRFPGKKWRTVGIPLEGIEELIPLDWNKKVMVERYE 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 445

A DNA sequence (GBSx0482) was identified in *S.agalactiae* <SEQ ID 1429> which encodes the amino acid sequence <SEQ ID 1430>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-546-

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18708 GB:U38906 ORF33 [Bacteriophage rlt]
Identities = 89/130 (68%), Positives = 106/130 (81%), Gaps = 1/130 (0%)

5 Query: 1 MTNFATDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETP-YFAT 59
M FAT DD+ +LWR L DE +RAE LLE VSD+LR EA KVG++L MI E P YFA+
Sbjct: 1 MNPFATVDDLTLMLWRPLKGDKEKRAEKLLEIVSDSLREEDKVGGRDLYAMIAEKPSYFAS 60

10 Query: 60 VLKSVTVDIVARTLMTATQGEPMQESQSQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKK 119
V+KSVTVDIVARTLMT+T EPM+Q ++SALGY+ SG+YLVPGGGLFIK+SEL RLGLKK
Sbjct: 61 VVKSVTVDIVARTLMTSTDQEPMTQTTESALGYSVSGSYLVPGGGLFIKNSLSRLGLKK 120

15 Query: 120 QRYGGIELYG 129
QR+G I+ YG
Sbjct: 121 QRFGVIDFYG 130

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1431> which encodes the amino acid sequence <SEQ ID 1432>. Analysis of this protein sequence reveals the following:

20 Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2061(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 116/138 (84%), Positives = 129/138 (93%)

30 Query: 3 NFATDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETPYFATVLK 62
NFATDDVILLWR LSVDE+KRA ALL+ VSDTLR+EA KVGK+LD+ +++ PYF V+K
Sbjct: 3 NFATDDVILLWRPLSVDELKRANALLKVSDTLRMEADKVGKDLDTMVDKPYFVNVIK 62

35 Query: 63 SVTVDIVARTLMTATQGEPMQESQSQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122
SVTVDIVARTLMT+T+GEPM+QESQSQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY
Sbjct: 63 SVTVDIVARTLMTSTRGEPMAQESQSQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122

40 Query: 123 GGIELYGEIERNNNSYFSR 140
GGIELYGEIER+NS FSR
Sbjct: 123 GGIELYGEIERDNSCFSR 140

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 446

A DNA sequence (GBSx0483) was identified in *S.agalactiae* <SEQ ID 1433> which encodes the amino acid sequence <SEQ ID 1434>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

50 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-547-

>GP:AAB18706 GB:U38906 Structural protein [Bacteriophage rlt]
 Identities = 132/296 (44%), Positives = 189/296 (63%), Gaps = 8/296 (2%)

Query: 5 IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEQKL 64
 + GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D +V E +K
 Sbjct: 3 LNKGTLDFTLVTDLISKVACKSSIALRLSAQKPIPFNGEKVFTFTMDSEIDVVAESGKKT 62

Query: 65 GNTAKVTSKIIPKLFVYQARMTDEFKYASEEKRLNFKHYADGFAKKMAEAFDIAIHG 124
 + + + P+K Y AR++DEF YAS+E+++N L+ + DGAFAK+A D+ A HG
 Sbjct: 63 HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

Query: 125 LEPRTMTDASFKATNSFDGVVTGNVIKYEADK--IDN--IDAAVTTIVANGNDVTGIAL 180
 + PR T ++ TN FD VT K EA + D N I+ AV + DVTGIA+
 Sbjct: 123 VNPRLGTASAVIGTNHFDKVTQ---KVEAPRGIADENGAIENAVELLTGVDADVTGIAI 179

Query: 181 SPQAGQDMSKRKDKFDNVMPYEPFRGQRPNSFFNMFLDINKTLTMKGGTAKDDHAIVGDF 240
 +P ++K+KD DN ++PE ++G P + +D+NKT++ T + D AI+GDF
 Sbjct: 180 NPSFRSALAKQKDLQDNALFPELKWGATPDITINGLPVDVNVKTVSDMSLTQR-DRAIIGDF 238

Query: 241 QNMFKGYAENIPMEIIEYGDGPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
 N FKWGYA+ +P+E+I+YGDGPD SG DLK YN++ +R E F+GWGILD F+RV
 Sbjct: 239 ANGFKWGYAKEVPLEVIQYGDGPDNSGLDLKGYNQVYIRAEFLGWGILDATKFARV 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1435> which encodes the amino acid
 sequence <SEQ ID 1436>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2772(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 133/298 (44%), Positives = 187/298 (62%), Gaps = 2/298 (0%)

Query: 1 MAESIKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEG 60
 M +LF LV++++KVKGHS+LAKLS Q PIPFNG ++F F LD + +V E
 Sbjct: 1 MGTETSKASLFDKHLVSDLINKVKGHSSLAKLSQKPIPFNGSKEFTFTLSDIDVVAEN 60

Query: 61 EQKLGNATKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFKHYADGFAKKMAEAFDIA 120
 +K + I P+K Y AR++DEF YA+EE+++ LK + +GFAKK+A D+
 Sbjct: 61 GKKTGGGLSLEPVTIVPIKVEYGARLSDEFYATEEEKIDILKAFNEGFAKKLARGIDLM 120

Query: 121 AIHGLEPRTMTDASFKATNSFDGVVTGNVIKYEADKIDNIDAAVTTIVANGNDVTGIAL 180
 A+HG+ PRT + TN FD VT V E++ D NI+AAV I + VTG+A+
 Sbjct: 121 AMHGINPRTKKASDVIGTNHFDKVTQVVKFTESEDADANIEAAVNLIQSGSEGVVTGLAM 180

Query: 181 SPQAGQDMSK-RKDKFDNVMPYEPFRGQRPNSFFNMFLDINKTLTMKGGTAKD-DHAIVG 238
 + ++K + MYPE +G P + + +N I+ A+ D I+G
 Sbjct: 181 DTEFSTALAKVTNGEMGPKMYPALAWGANEDSINCLKSSVNTTVGAGADEAESKDLVIIG 240

Query: 239 DFQNMFKWGYAENIPMEIIEYGDGPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
 DF++MFKWGYA+ IPMEII+YGDGPD SG+DLK YN+I LR EA+IGWGILD K+F+RV
 Sbjct: 241 DFESMFKWGYAKQIPMEIIKYGDGPDNSGDLKGYNQIYLRAEAYIGWGILDAKSFARV 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 447

A DNA sequence (GBSx0484) was identified in *S.agalactiae* <SEQ ID 1437> which encodes the amino acid sequence <SEQ ID 1438>. Analysis of this protein sequence reveals the following:

-548-

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2224(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9659> which encodes amino acid sequence <SEQ ID 9660>
10 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18705 GB:U38906 CRF30 [Bacteriophage rlt]
Identities = 64/158 (40%), Positives = 101/158 (63%), Gaps = 8/158 (5%)

15 Query: 43 MSEFKVIETQEELDTIVKARIARERE----KYQDYDQLKTRVEELETENSSLQTALNDAK 98
 MSE + +TQEEL+ I++ R+AR++E + D+D+LKT++ LE +N++ Q + ++K
 Sbjct: 1 MSENNLPKTQEELNQIETRLARQKETIEANFADYDELKTKIAALEADNTAYQATIEESK 60

20 Query: 99 SNTDSYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDDEDGLKVDARLAS 158
 S + ++ E QI+GY+ L+ +A++ GLP+DLA+RL GDDE+ LK DAER +
 Sbjct: 61 S---WEQEKADYEQKISGYKTTQLKQSIKAGLPLDLADRLSGDDEESLKADAERFSG 116

25 Query: 159 FIKPSQPQPPTKSNEPIITDQKEAGWIEMARNLVNKG 196
 FIKP P P K EP + D K+ + ++ L +GE
 Sbjct: 117 FIKPKTPAPLKDVEPNLGDGKDGAIRKLVNGLKTEGE 154

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1439> which encodes the amino acid
sequence <SEQ ID 1440>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3476(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 128/149 (85%), Positives = 136/149 (90%)

40 Query: 43 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 102
 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD
 Sbjct: 1 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 60

45 Query: 103 SYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDDEDGLKVDARLASFIKP 162
 SYTE+I+TL+NQIA YE ANLRTKVALQYGLPIDLA+RLQGDDDEDGLKVDARLASFIKP
 Sbjct: 61 SYTEEISTLKNQIADYETANLRTKVALQYGLPIDLADRLQGDDDEDGLKVDARLASFIKP 120

50 Query: 163 SQPQPPTKSNEPIITDQKEAGWIEMARNL 191
 SQPQPP KSNP I +A + + + L
 Sbjct: 121 SQPQPPAKSNEPNIDSNADANYRALVQGL 149

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 448

55 A DNA sequence (GBSx0485) was identified in *S.agalactiae* <SEQ ID 1441> which encodes the amino
acid sequence <SEQ ID 1442>. Analysis of this protein sequence reveals the following:

Possible site: 56

-549-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:AAB18704 GB:U38906 ORF29 [Bacteriophage rlt]
 Identities = 322/461 (69%), Positives = 383/461 (82%)

15 Query: 8 KLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNLWVHQKYG 67
 + GNQ PTQSV L F +T EAI Y+K+ CY WQ+N+L +MAI+ED LW HQK+G
 Sbjct: 6 RFGNQYPTQSVILPFTETKYQEAIEIYEKSKHECYPWQKNLLKEVMAIDEDGLWTHQKFG 65

20 Query: 68 YAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGYVDGEDFI 127
 Y+IPRRNGKTE+VYILELW+L +GL ILHTAHRISTSHSS+EK+KKYLE SGYV+GEDF
 Sbjct: 66 YSIPRRNGKTEIVYILELWLSLVQGLSILHTAHRISTSHSSYEKLLKYLEDSGYVEGEDFK 125

25 Query: 128 SNKAKQERIEFKSSGSVIQPRTRTSNGGLGEGFDLLIIDEAQEYTAEQESALKYTVTDS 187
 S KAKQER+E SG VIQPRTRTS+GGLGEGFD+L+IDEAQEYT EQESALKYTVTDS
 Sbjct: 126 SIKAKQERLELIESGGVIQPRTRTSNGGLGEGFDILVIDEAQEYTTTEQESALKYTVTDS 185

30 Query: 188 DNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRLYSGWAEWSVDEMOPHDKVSWYVANP 247
 DNPMTIMCGTPPT VS+GTVF +YR + G +YSGWAEWSV++++ IHDV++WY +NP
 Sbjct: 186 DNPMTIMCGTPPTPVSSGTVFTNYRDNTIAGKAKYSGWAEWSVEDVKDIHDVEAWNSNP 245

35 Query: 248 SMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVPELKSKLF 307
 SMGYHLNERKIEAELGED++DHN+ORLGYWP +NQKSVISE+EW LKV ++P +K KLF
 Sbjct: 246 SMGYHLNERKIEAELGEDKLDHNVQRLGYWPKYNQKSVISEQEWNAKLVNRLPVIKGLF 305

40 Query: 308 VGIKFGQDGNVSLSIAARASENKVFVEAIDCLSVRNGTQWIIINFLKSADIKVVVDGAS 367
 VGIK+G DG NV++SIA + KVFVE IDC S+RNG QWIIINFLK AD+ KVV+DG S
 Sbjct: 306 VGIKYNDGANVAMSIKVTLSGKVFVETIDCQSIRNGNQWIIINFLKADVEKVVVDGQS 365

45 Query: 368 GOELLAQEMREHGLKKPELKPVAEIIITANTMWEQGIMQETICHNDQPSLTAVVTNCKEKRQ 427
 GQ +L EM++ LK+P LP V EII AN++WEQGI Q+ CH+ QPSL+ VVTNC+KR
 Sbjct: 366 GQSILITSEMKDFKLKEPILPTVKEIINANSLWEQGIFQKNFCHSGQPSLSTVVTNCDKRN 425

50 Query: 428 IGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQR 468
 IG++GGFGYKS +DD DISLMDSALLAHW C KPK+KQ+
 Sbjct: 426 IGTSGGFGYKSQFDDMDISLMDSALLAHWACSNNKPKKKQQ 466

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1443> which encodes the amino acid
 sequence <SEQ ID 1444>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3133(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 437/471 (92%), Positives = 459/471 (96%)

60 Query: 1 MVTKTAKLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNL 60
 MVTKTAKLGNQRPTQSVNLHFAK+LAHEAINYYKKTGLSCY WQ NMLIP+MAI+E+ L
 Sbjct: 6 MVTKTAKLGNQRPTQSVNLHFAKSLAHEAINYYKKTGLSCYPWQVNMMLIPIMAIIDENGL 65

60 Query: 61 WVHQKYGYAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGY 120
 WVHQKYGYAIPRRNGKTEVVYI++LWALHKGLKILHTAHRISTSH+SFEKVKKYLEMSGY
 Sbjct: 66 WVHQKYGYAIPRRNGKTEVVYIVQLWALHKGLKILHTAHRISTSHASFEKVKKYLEMSGY 125

-550-

Query: 121 VDGEDFISNKAKGQERIEFKSSGSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESAL 180
 VDGEDFISNKAKGQERIEFK+SG+VIQFRTRTSNGGLGEGFDLLIIDEAQEYT+EQESAL
 Sbjct: 126 VDGEDFISNKAKGQERIEFKASGAVIQFRTRTSNGGLGEGFDLLIIDEAQEYTSEQESAL 185

5 Query: 181 KYTVTDSNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSWAESVDEMQUIHDVK 240
 KYTVTDSNPMTIMCGTPPTMVSTGTVFE+YRK+CLKG++RYSWAESV EM I+DV
 Sbjct: 186 KYTVTDSNPMTIMCGTPPTMVSTGTVFEAYRKDCLKGNKRYSGWAESVPEMVKINDVS 245

10 Query: 241 SWYVANPSMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 300
 SWY++NPSMG+HLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP
 Sbjct: 246 SWYISNPSMGFHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 305

15 Query: 301 ELKSKLFGVGIKFGQDGNVSLSLAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAK 360
 ELKSKLFGVGIKFGQDGNVSLSLAAR SENKVFE IDCLSVRNGTQWIINFLKSADIAK
 Sbjct: 306 ELKSKLFGVGIKFGQDGNVSLSLAARTSENKVFEVETIDCLSVRNGTQWIINFLKSADIAK 365

20 Query: 361 VVVDGASGQELLAQEMREHGLKKPELPKVAEIIITANTMWEQGIMQETICHNDQPSLTAVV 420
 VV+DGASGQELLAQEM++ GLKKPELPKVAEIIITAN MWEQGIMQETICH+DQPSLTAVV
 Sbjct: 366 VVIDGASGQELLAQEMKDQGLKKPELPKVAEIIITANMMWEQGIMQETICHSDQPSLTAVV 425

Query: 421 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC 471
 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC
 Sbjct: 426 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC 476

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 449

A DNA sequence (GBSx0486) was identified in *S.agalactiae* <SEQ ID 1445> which encodes the amino acid sequence <SEQ ID 1446>. Analysis of this protein sequence reveals the following:

30 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.2745(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 450

A DNA sequence (GBSx0487) was identified in *S.agalactiae* <SEQ ID 1447> which encodes the amino acid sequence <SEQ ID 1448>. Analysis of this protein sequence reveals the following:

45 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.2568(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18703 GB:U38906 ORF28 [Bacteriophage r1t]

-551-

Identities = 124/250 (49%), Positives = 164/250 (65%), Gaps = 3/250 (1%)

Query: 2 VDDVLPKLLKSVQQDFEKHFQKSEVVAKAFELQAKKATYKTVNEFAVEVGRLLSLALAN 61
 ++D+LP LL+ + QDF++ S+ + ++ L+ KKATY NEF VEVG++LS L
 5 Sbjet: 1 MEDILPPLLEKINQDFDERAANSKKLKQSMELLLTKKATYIQANFEGVEVGQILSDVLGT 60

Query: 62 SVISDELPGKMYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPPLNQ 121
 V D LPDGKMY+NIA+RL+N L+ N+ LIS Y+ DVQ LN+ A LK Q P LNQ
 10 Sbjet: 61 HVTVDVLPDGKMYFNIAADRLLSILKKNFDLISGYSTDVQSELNQLAGFKLKSQVPELNQ 120

Query: 122 DKIDGLVNLASEPVFDDVKWLLDEPIVNFSSQIVDDCIRANADFHFKTGLKPTIERIST 181
 D+IDG+VNR++SE F+ + WLL EPIV FSQS+VDD ++ N DF K GLKP I R
 15 Sbjet: 121 DRIDGIVNRISSEDDFEKILWLLKEPIVTFSSQSVVDDTLKKNIDFQAKAGLKP KIVRKL 180

Query: 182 GKCCDWCDRLAGRYVYHEEPKDFYKRQHQCQVIDYHPK--NGKRQNSWSKKWTKETTDI 239
 GK CDWC LAG Y Y P D Y RH+ C+C ++Y P+ + KRQ+ WSK W D
 20 Sbjet: 181 GKACDWCRNLASGYDYPNVPSDVYHRHERCRCTVEYDPRDIDKKRQDVWSKNWVDPDKDA 240

Query: 240 -LERRKQMN 248
 + RK +N+
 20 Sbjet: 241 KIAERKNLNL 250

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1449> which encodes the amino acid sequence <SEQ ID 1450>. Analysis of this protein sequence reveals the following:

25 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3099(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 169/261 (64%), Positives = 207/261 (78%), Gaps = 2/261 (0%)

35 Query: 1 MVDDVLPKLLKSVQQDFEKHFQKSEVVAKAFELQAKKATYKTVNEFAVEVGRLLSLALA 60
 MVDDVLPKLLKSV+QDFEK+FG+S+V V KAFELQAKK TYKTVNEFA+EVGRLLSLAL
 Sbjet: 1 MVDDVLPKLLKSVRQDFEKYFGESDVVTKAFELQAKKVITYKTVNEFAIEVGRLLSLALT 60

40 Query: 61 NSVISDELPGKMYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPPLN 120
 SV SD+LPDGKMYNIA RL+++T+ NYKLIS YAGDVQ+ LN+ A+I LK+QRPPLN
 Sbjet: 61 GSVSSDKLPDGKMYNIAKRLLEDTMGRNYKLISGYAGDVQRIILNENAQIGLKVQRPPPLN 120

45 Query: 121 QDKIDGLVNLASEPVFDDVKWLLDEPIVNFSSQIVDDCIRANADFHFKTGLKPTIERIS 180
 +DKI+G+VNRL SE FDDVKWL EPIVNFSSQIVDD I+ANAD +KTG+ P + R
 Sbjet: 121 RDKINGMVNRLDSENTFDDVKWLFGEPIVNFSSQIVDDTIKANADLQYKTMTPQVVRTE 180

50 Query: 181 TGKCCDWCDRLAGRYVYHEEPKDFYKRQHQCQVIDYHPKNGKRQNSWSKKWTK--ETTD 238
 +G CC+WC + G Y Y + PKD ++RHQ C+C +DY PKNGK Q++WSK W K +T +
 Sbjet: 181 SGNCCEWCREVVGCTYSYPKPKDVWRRHQRCTLDYDPKNGKVQSAWSKIWRKKEKTQE 240

Query: 239 ILERRKQMNIDIRDNNRKS 259
 +ER ++ + K+DI
 55 Sbjet: 241 SIERVEKFESALVESIKNDI 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 451

A DNA sequence (GBSx0488) was identified in *S.agalactiae* <SEQ ID 1451> which encodes the amino acid sequence <SEQ ID 1452>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 58
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.38    Transmembrane    93 - 109 ( 93 - 110)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAC39307 GB:AF022773 ORF5 [Lactococcus bacteriophage phi31]
      Identities = 271/410 (66%), Positives = 326/410 (79%), Gaps = 2/410 (0%)

Query: 1   MNYMGMYLQRLALFKTGVDKRYRYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
      M   G+GYL+ KL++ K   + RY  YAM   D   + I +P   + + YRS++ W AKGVD
20  Sbjct: 1   MTEKGIGYLRFKLSVHKRAEMRYEQYAMKHVDRFKGITIPQALSQQYRSILGWCAKGVD 60

Query: 61   SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
      SLADR+IFREF NDDF   EIF+ NNPDIFFD+A+ SALIASC F+YI   G+ D++ ++Q
25  Sbjct: 61   SLADRLIFREFENDDFTVNEIFEENNPDIFFDSAVLSALIASCSFTIYISKGENDAV-RLQ 119

Query: 121  VIEASKATGILDPTTFLLTEGYAVLESNSNENPTLEAYFTGEKTWYYPKDEKP-YSIDNS 179
      VIEA  ATGI+DP T LLTEGYAVLE D N N  LEA+F   ++T YY +D +   SI N
30  Sbjct: 120 VIEAVNATGIIDPITGLLTEGYAVLERDENNNVVLEAHFLPDRTDYYYRDSRNNISIANP 179

Query: 180  TGHPLLVPIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMD 239
      TGHPLLVPIHRPDAVRPFGRSRIT++GMY Q  AKRTLERA+VTAEFYSFPQKYV G+
35  Sbjct: 180 TGHPLLVPIIHRPDAVRPFGRSRITRSGMYWQSNKRTLERAADVTAEFYSFPQKYVTGLS 239

Query: 240  PDAEPMEKWRATVSTLLEISKDEDEGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGGLT/L 299
      DAEPME W+ATVS++L+ +KDEGDKPT+GQFT  SM+PF + L+  A+  FAG +GLTL
40  Sbjct: 240 DDAEPMETWKAIVSSMLQFTKDEGDKPTLGQFTQPSMSPFTEQLRTAAAGFAGETGLTL 299

Query: 300  DDLGFPSDNPNSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFM 359
      DDLGF SDNPNSSVEAIKA+HENLR AGRKAQRS  +G LNVAY+A CLRDD PYLR QF
45  Sbjct: 300 DDLGFVSDNPNSSVEAIKASHENLRLAGRKAQRSLGAGLLNVAYLAACLRDDVPYLRQFS 359

Query: 360  DTEIKWEPLFEADANMLTLVGDAIKLNQAIPGFMADAVIRDLTGKVGSD 409
      T+  KWEPLFEADA+ML+L+GDGAIKLNQAIP F++ D IRDLTG+KG++
50  Sbjct: 360 KTKPKWEPLFEADASMLSLIGDGAIKLNQAIPFINKDTIRDLTGKIGAE 409

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1453> which encodes the amino acid sequence <SEQ ID 1454>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 58
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.38    Transmembrane    93 - 109 ( 93 - 110)

55  ----- Final Results -----
      bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

      Identities = 395/422 (93%), Positives = 407/422 (95%)

60  Query: 1   MNYMGMYLQRLALFKTGVDKRYRYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
      MNYMGMYL+RKLALFKTGVDKRYRYAMDDRD+TRSIVMP+NVREMYRSV+EWTAKGVD

```

-553-

Sbjct: 1 MNYMGMGYLRRKLALFKTGVDKRYRYAMDDRDTRSIWPNVREMYRSVLEWTAGVD 60

Query: 61 SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
SLADRIIFREF NDDFNWEIFKANNPDIFFDTAIQSALIASCCFVYIMPG ED LPKMQ

5 Sbjct: 61 SLADRIIFREFNDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGAEDGLPKMQ 120

Query: 121 VIEASKATGILDPTTFLTEGYAVLESNSNPTLEAYFTGEKTYWYYPKDEKPYSDNST 180
VIEASKATGILDPTTFLTEGYA+LESNSN NPTLEAYFT + WYYPK KPY+I N T

10 Sbjct: 121 VIEASKATGILDPTTFLTEGYAILESNSNPTLEAYFTDKDIWYYPKKGKPYNIKNPT 180

Query: 181 GHPLLPVVIHRPDVAVRPFGSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240
GHPLLPV+IHRPDVAVRPFGSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP

Sbjct: 181 GHPLLPVVIHRPDVAVRPFGSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240

15 Query: 241 DAEPMEKWRATVSTLLEISKDEDDGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTLD 300
DAEPMEKWRATVSTLLEISKDEDDGDKPTVGQFTTASMAPFM+HLKMYASLFAGGSGLTLD

Sbjct: 241 DAEPMEKWRATVSTLLEISKDEDDGDKPTVGQFTTASMAPMEHLKMYASLFAGGSGLTLD 300

20 Query: 301 DLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDPFYLRNQFMD 360
DLGFPSDNPSSVE+IKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRD+FPYLRNQFMD

Sbjct: 301 DLGFPSDNPSSVESIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDEFPYLRNQFMD 360

25 Query: 361 TEIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGKGSNDNPIPKATEVTT 420
T IKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGKVG+D PTP TEVTT

Sbjct: 361 TVIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGKVGADKPIPAITEVTT 420

Query: 421 DG 422
DG

30 Sbjct: 421 DG 422

SEQ ID 1452 (GBS364) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 6; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 11; MW 75kDa).

GBS364-GST was purified as shown in Figure 216, lane 10.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 452

A DNA sequence (GBSx0489) was identified in *S.galactiae* <SEQ ID 1455> which encodes the amino acid sequence <SEQ ID 1456>. Analysis of this protein sequence reveals the following:

40 Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.4063(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1457> which encodes the amino acid sequence <SEQ ID 1458>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4120(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-554-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 101/118 (85%), Positives = 110/118 (92%)

5 Query: 1 MKKKCLICKKTFQAKTNRSLYCSEECRKKGIREKQKRLMKQKRADKKKEKIKVLNTNADV 60
 +KKKCLICKK FQAKTNR+LYCSEECRKKG REKQKRLMKQKRA+++KEK KVLN N DV
 Sbjet: 1 LKKKCLICKKNFQAKTNRSLYCSEECRKKGNREKQKRLMKQKRAEQRKEKKKVLNPNTDV 60

10 Query: 61 TEKPKKIRNLVQHYKKLKREILDNESEFGFTGIALVEGIDIHEENFVDLVMQKIKEQQ 118
 TEKPKKIRNL QHYKKLK+EIL NESEFGFTGI L+EGID+HEENFVDLVMQKIKEQ+
 Sbjet: 61 TEKPKKIRNLQHYKKLKKEILANESSEFGFTGITLEGIDVHEENFVDLVMQKIKEQK 118

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 453

A DNA sequence (GBSx0490) was identified in *S.agalactiae* <SEQ ID 1459> which encodes the amino acid sequence <SEQ ID 1460>. Analysis of this protein sequence reveals the following:

Possible site: 19
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0633(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC39305 GB:AF022773 ORF3 [Lactococcus bacteriophage phi31]
 Identities = 75/109 (68%), Positives = 87/109 (79%), Gaps = 1/109 (0%)

30 Query: 29 LRADKKGTHRVAFEKNRRLKTAHLGICGPRVDKSLKYPHPLSAAIDHIVPIAKGGHP 88
 LRAD+ G HRVAF+KN++ LLKT + CGICG+P+DK LK P PLS +DHI+PI KGGHP
 Sbjet: 3 LRADRTGAHRVAFDKNRKILLKTQNTCGICGKPIDKRLKAPDPLSPVVDHIIPINKGGHP 62

35 Query: 89 SSIDNLQLTHWQCNROKSDKLFINQTAVRATVVGNRNLPQSRDWSSSYAS 137
 S++DNLQL HW CNROKSDKLF N V+GNRNLPQSRDWSSY S
 Sbjet: 63 SAMDNLQLAHWTCNRQKSDKLF-NVKQEEPVLGNRNLPQSRDWSSSYVS 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1461> which encodes the amino acid
 40 sequence <SEQ ID 1462>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.4185(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 88/112 (78%), Positives = 102/112 (90%)

50 Query: 28 KLRADKKGTHRVAFEKNRRLKTAHLGICGPRVDKSLKYPHPLSAAIDHIVPIAKGGH 87
 +LRADKKGTHRVAF++NK++LLK A +CGICG+PVDKSLKYPHPLSAAIDHIVPIAKGGH
 Sbjet: 3 QLRADKKGTHRVAFDRNKKLLKAATVCGICGKPVVDKSLKYPHPLSAAIDHIVPIAKGGH 62

55 Query: 88 PSSIDNLQLTHWQCNROKSDKLFINQTAVRATVVGNRNLPQSRDWSSSYASKE 139
 PS+++NLQLTHWQCNROKSDKLF NQ + +GNRNLPQSRDWSS+A K+
 Sbjet: 63 PSALENLQLTHWQCNROKSDKLFANQASNEPKTIGNRNLPQSRDWSSFAFKK 114

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 454

- 5 A DNA sequence (GBSx0491) was identified in *S.agalactiae* <SEQ ID 1463> which encodes the amino acid sequence <SEQ ID 1464>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 455

A DNA sequence (GBSx0492) was identified in *S.agalactiae* <SEQ ID 1465> which encodes the amino acid sequence <SEQ ID 1466>. Analysis of this protein sequence reveals the following:

20 Possible site: 28
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2907(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AAF43508 GB:AF145054 ORF15 [Streptococcus thermophilus
bacteriophage 7201]
Identities = 61/187 (32%), Positives = 90/187 (47%), Gaps = 31/187 (16%)
Query: 1 MNIEEAKKLIDKQSIGKGGVDIPVVKTHIVKVLDDQIDQPQPEVPRFVADWYEKHKDSL 60
MN +EA K I K+ + L D I +P VP++VADWYE+HKD
35 Sbjct: 1 MNRDEAVKKIAKEGY-----ISIEHAEDLYDSIIT-KPVVPQYVADWYEEHKDEF 49
Query: 61 ECDL-----YLYHMSIY--DEEVEKDDFFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYT 112
+L + H++ Y +E DF W +KN + L+NMHQFGY ++KEK YT
Sbjct: 50 YLNLHRVVRDFFFEHLNAYYFNENPIDYDFACWYNTKNAIQILVNMHQFGYEVKKEKRYT 109
40 Query: 113 VEIPN--PNERQLSFVLMRQLSGNVSIVMHRDNLDLLKTDNDLQLTESEIRKDFDWAQ 170
V I N E L++ R+ + RDN D +T + T E+ ++ + W
Sbjct: 110 VRIRNLDDDEETYLNYDKFRE-----TWVFYSRDNTDRFRTIH----THKEL-EEGGFGWV 159
45 Query: 171 FREEVVE 177
F E +E
Sbjct: 160 FDCEGIE 166

- 50 A related GBS nucleic acid sequence <SEQ ID 10927> which encodes amino acid sequence <SEQ ID 10928> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1467> which encodes the amino acid sequence <SEQ ID 1468>. Analysis of this protein sequence reveals the following:

Possible site: 21

-556-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3815(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 70/180 (38%), Positives = 98/180 (53%), Gaps = 30/180 (16%)

Query: 1 MNIEEAKKLIDKQSI-GKGGVGDIPVVKTHIVKVLDDQIDQPQPEVPRFVADWYEEKHKS 59
 MNIEEAK+L+D GK V+K V+ ++DQ++QP+PEVP+ VADW E+ K+
 Sbjct: 1 MNIEEAKELVDNSKFGKTS----SVIKAE-VRDIIDQLNQPKPEVPQCVADWIEECKEE 55

15 Query: 60 LECDLYLYHMSIYDEEVEKDDFFYYWQTSKNPVYTLINMHQFGYTIQKEKLYTVEIPN-- 117
 DL L ++ + W+ S + GYT++KEKLYTV++PN
 Sbjct: 56 ---DLTL--KGLFSNSDMPAKIFDWIFGSDENCRMLMAEAWINGYTVEKEKLYTVDLPNGQ 110

20 Query: 118 PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQFREEVVE 177
 P R ++ + Q L T+N ++LTESEIRKDF+WAWQF EEV E
 Sbjct: 111 PLVRGINTLYFSON-----LATEN-VKLTESEIRKDFEWAWQFAEEVTE 153

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 **Example 456**

A DNA sequence (GBSx0493) was identified in *S.agalactiae* <SEQ ID 1469> which encodes the amino acid sequence <SEQ ID 1470>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.5365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 **Example 457**

A DNA sequence (GBSx0494) was identified in *S.agalactiae* <SEQ ID 1471> which encodes the amino acid sequence <SEQ ID 1472>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -8.55 Transmembrane 34 - 50 (31 - 54)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9657> which encodes amino acid sequence <SEQ ID 9658> was also identified.

-557-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1473> which encodes the amino acid sequence <SEQ ID 1474>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-11.25    Transmembrane    26 - 42 ( 20 - 49)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

15      Identities = 56/89 (62%), Positives = 71/89 (78%)

      Query: 8  MTEQQMIDCLLYELAKKDKLNIRRNNIITFLSIVLMAISILNVALQDHYKSQITELRTQL 67
                MTE+QMIDCLLYEL KKDK   +++ II  L+++L+ +S L V+L+ +Y+ QI  LRTQL
20      Sbjct: 1  MTEQQMIDCLLYELVKKDKAIKKKSIIIAALTVMILIVVSLGCVSLKSYIEPQIYGLRTQL 60

      Query: 68  SRTQKQLKRASDDRARQTKRIAELTGNGG 96
                SRTQKQLKRAS+  RQTKRIA+LT NGG
      Sbjct: 61  SRTQKQLKRASEQNQRQTKRIADLTNNGG 89

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 458

A DNA sequence (GBSx0495) was identified in *S.agalactiae* <SEQ ID 1475> which encodes the amino acid sequence <SEQ ID 1476>. Analysis of this protein sequence reveals the following:

```

30      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.2040(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 459

A DNA sequence (GBSx0496) was identified in *S.agalactiae* <SEQ ID 1477> which encodes the amino acid sequence <SEQ ID 1478>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 34
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.3044(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-558-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAD37108 GB:AF109874 unknown [Bacteriophage Tuc2009]
 Identities = 50/143 (34%), Positives = 67/143 (45%), Gaps = 29/143 (20%)

Query: 1 MIPNFRFNFKETKKM-YG-VDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDNFDDYIL 58
 MIP RA++K+ ++M YG V+ F+ S+ YR HF +D
 10 Sbjct: 1 MIPKLRWDKQDERMSYGEVEYFDDSI--YRFD-----HFCTGADEDVEF 44

Query: 59 MQSTGMFDKNGVEIFDGDIVLTTRL-----IDY-TYKNFKGVVKMLEGRWLIDTGKDA 110
 MQSTG+ DKNQVEI++GDI+ + I Y Y G + EG L +
 15 Sbjct: 45 MQSTGIKDKNGVEIYEGDILKLHAIFLAPDDKIGYLEYSPKYGYSIICEGNRLY---RQE 101

Query: 111 VGLWTEVDENEAGNIYQNSSELL 133
 T E IGNIY+N ELL
 20 Sbjct: 102 YWASTNKLNYEVIGNIYENPELL 124

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1479> which encodes the amino acid sequence <SEQ ID 1480>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 44/52 (84%), Positives = 47/52 (89%)

Query: 1 MIPNFRFNFKETKKMYGVDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDN 52
 MIPNFR FNK+TKMY +DGF+ S RKIYRCSLADDEFRCGRLETFHFVEDN
 35 Sbjct: 1 MIPNFRGFNFKTKKMYSIDGFKSSERKIYRCSLADDEFRCGRLETFHFVEDN 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 460

40 A DNA sequence (GBSx0497) was identified in *S.agalactiae* <SEQ ID 1481> which encodes the amino acid sequence <SEQ ID 1482>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9655> which encodes amino acid sequence <SEQ ID 9656> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-559-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 461

A DNA sequence (GBSx0498) was identified in *S.agalactiae* <SEQ ID 1483> which encodes the amino acid sequence <SEQ ID 1484>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5189 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9653> which encodes amino acid sequence <SEQ ID 9654> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF43503 GB:AF145054 ORF10 [Streptococcus thermophilus
bacteriophage 7201]
Identities = 92/147 (62%), Positives = 121/147 (81%)
Query: 15 IEPKPQTRPKFSKFGTYEDPKMKRWRKEVSGWIEKNYDGPFFDDCIKVEVTFYMKAPKTL 74
IEPKPQTRP+FSKFGTYEDPKMK WR+E S IE+ YDG FF I V+VTFYMKAP ++
Sbjct: 7 IEPKPQTRPRFSKFGTYEDPKMKAWRRECSRLIEQEYDGGQFFYGPISVDVTFYMKAPLSV 66
Query: 75 SKEPTQRSKGGKTIQIVQNFVRELIWHAKKPDIDNLIKAVFDSISDAGYDRIQKSGIVWSD 134
SK+PT +++ KT ++ F+ E +WH++KPDIDNLIKA+FDsis AGY+++ K GIVW+D
Sbjct: 67 SKKPTPKARAKTWDAFKKFMAERLWHSRKPDIDNLIKALFDSISDAGYNKVDKKGIVWTD 126
Query: 135 DNIVCDLRAKKKYQNPRIKVRIEEID 161
D+IVC L A+K+YS+NERI+ I+E++
Sbjct: 127 DSTVCKLSAQKRYSENPRIEFEIKELE 153

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 462

A DNA sequence (GBSx0499) was identified in *S.agalactiae* <SEQ ID 1485> which encodes the amino acid sequence <SEQ ID 1486>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.4007 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-560-

Example 463

A DNA sequence (GBSx0500) was identified in *S.agalactiae* <SEQ ID 1487> which encodes the amino acid sequence <SEQ ID 1488>. This protein is predicted to be pXO1-07. Analysis of this protein sequence reveals the following:

```

5   Possible site: 26
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3664 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
    Identities = 146/373 (39%), Positives = 216/373 (57%), Gaps = 18/373 (4%)

    Query: 35  LYDKVYRKDILKVAWFYVVRNKGSGAGIDFTTIEEIEAYGVQKFLDEIEDQLRNKKYQPKA 94
               L DK+  ++ +  A+  VK NKGSGAGID TIEE++ Y Q +  ++ ++ +KY+P+
20  Sbjct: 4    LLDKILSRENMLEAYNQVSKNKGSGAGIDGMTIEEMDNYLRQNWRLTKELIKQRKYKQPQ 62

    Query: 95  VKRVYIPKANGKKRPLGIPTVRDRVVQTAVKIVIEPIFEADFQEFSGFRPKRSANQAIR 154
               V +V IPK +G  R LGIPTV DR++Q A+  V+ PI E  F + SYGFRP RS  +AI
    Sbjct: 63  VLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKAIM 122

25  Query: 155 EIYKYLNYGCEWVIDADLKGYFDTI PHDKLLLLVKERVTDKSIKLLSLWLEAGIMEDNQ 214
               ++ +YLN G EW++D DL+ +FDT+P D+L+ LV  + D  L+  +L +G++ + Q
    Sbjct: 123 KLELYLNDGYEWIVDIDLEKFFDTPVQDRLMSLVHNIIEDGDTESLIRKYLHSGVIINGQ 182

    Query: 215 VRSNLTGPQGGVISPLLANIYLNALDRYWKNNRLEGRGHDAHLIRYADDFVI-LCSNNP 273
               ++GTPQGG +SPIL+NI LN LD+  LE RG  +RYADD VI + S
30  Sbjct: 183 RYKTLVGTGPQGGNLSPLLSNIMLNELDK-----ELEKRG--LRFVRYADDCVITVGSEAA 235

    Query: 274 KKYVYQYAKQRI--DKLGLTLNBEKTRIVHATEGFDLGYTLRKS KSHKSGKYKTYVPSR 331
               K  Y+  R  +LGL +N  KT+I  E  +LG+  KS  +  P +
35  Sbjct: 236 AKRVMYSVSRFIEKRLGLKVNMTKTKITRPRE-LKYLGFGEWKSSDGWKS-----PHQ 288

    Query: 332 KSMKSIKGVKVDVIQTGQHLNLPDVMERLNPMLRGWANYFKAGNSKHFKSIDNYVIYNL 391
               S++  K K+K + Q  ++L  +E+LN  +RGW NYF  GN K  SID +  L
40  Sbjct: 289 DSVRRFKLKLKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRLTRL 348

    Query: 392 TIMLRKHKHKKSGK 404
               +++ K+ KK  +
    Sbjct: 349 RMIIWKQWKKKSR 361

```

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 464

50 A DNA sequence (GBSx0501) was identified in *S.agalactiae* <SEQ ID 1489> which encodes the amino acid sequence <SEQ ID 1490>. Analysis of this protein sequence reveals the following:

```

    Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
55      bacterial cytoplasm --- Certainty=0.3833 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-561-

A related GBS nucleic acid sequence <SEQ ID 9651> which encodes amino acid sequence <SEQ ID 9652> was also identified.

A further related DNA sequence (GBSx2517) was identified in *S.agalactiae* <SEQ ID 7217> which encodes the amino acid sequence <SEQ ID 7218>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1491> which encodes the amino acid sequence <SEQ ID 1492>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 27
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
20  bacterial cytoplasm --- Certainty=0.2299(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25  Identities = 113/163 (69%), Positives = 128/163 (78%), Gaps = 25/163 (15%)

Query: 1  MINNIVLVGRMTKDAELRYTPSNQAVATFSLAVNRNFKNQSGEREADFINCVIWRQQAEN 60
      MINN+VLVGRMTKDAELRYTPS  AVATF+LAVNR  FK+Q+GEREADFINCVIWRQ  AEN
Sbjct: 1  MINNVVLVGRMTKDAELRYTPSQVAVATFTLAVNRTFKSQNGEREADFINCVIWRQPAEN 60

30  Query: 61  LANWAKKGALVGITGRIQTRNYENQQGQRIYVTEVVAENFOLLESRSNQ-----Q 111
      LANWAKKGAL+G+TGRIQTRNYENQQGQR+YVTEVVA+NFQ+LESR  +++
Sbjct: 61  LANWAKKGALIGVTGRIQTRNYENQQGQRVYVTEVVADNFQMLSRATREGGSTGSGFNGG 120

Query: 112  TNQSGNSSNSY-----FGNANKMDISDDDLPF 138
35  N + +SSNSY                      FGN+N MDISDDDLPF
Sbjct: 121  FNNNTSSNSYSAPAQQTENFGRDDSPFGNSNPMDISDDDLPF 163

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 465

A DNA sequence (GBSx0502) was identified in *S.agalactiae* <SEQ ID 1493> which encodes the amino acid sequence <SEQ ID 1494>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
45  >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.33    Transmembrane    17 - 33 ( 17 - 33)

   ----- Final Results -----
50  bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-562-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 466

A DNA sequence (GBSx0503) was identified in *S.agalactiae* <SEQ ID 1495> which encodes the amino acid sequence <SEQ ID 1496>. This protein is predicted to be p22 erf-like protein. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA97824 GB:AB044554 orf 17 [Staphylococcus aureus prophage

phiPV83]

Identities = 93/183 (50%), Positives = 120/183 (64%), Gaps = 5/183 (2%)

Query: 1 MRKSESITEYAKAFCKAQLEVKQPLKDKDNPFKSKYVPLENVTEAITTAFANNGISFSQ 60
 M KSE++ E KA + EVKQPLKDK+NPFFKSKYVPLENV EAI A +G+S++Q
 Sbjct: 1 MNKSETVVEINKAMVAFRKEVKQPLKDKNNPFKSKYVPLENVVEAIDEAATPHGLSYTQ 60

Query: 61 DPTTNTENGYIDVATLVMHTSGEWVEYGPLSVKPTKNDVQGAGSAITYAKRYALSAIFGI 120
 N +G + VAT++MH SGE++EY P+ + KN QGAGS I+Y KRY+LSAIFGI
 Sbjct: 61 W-ALNDVDGRVGVATMLMHESGEYIEYDPVFMNAEKNTPOGAGSLISYLKRYSLSAIFGI 119

Query: 121 TSDQDDDGNEGSKPNNRSQSPKATTKKTKGTGYQTPKISNIQIETYKSDLNDIAKATNQ 180
 TSDQDDDGNE S NN +PK T +TQ +T I ++ ++ + K QN
 Sbjct: 120 TSDQDDDGNEASGKNN--NPKQQT-RTQWASSETIGILRKEVISFTKLIKGTDKAPQN 175

Query: 181 VEE 183

+ E

Sbjct: 176 IVE 178

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 467

A DNA sequence (GBSx0504) was identified in *S.agalactiae* <SEQ ID 1497> which encodes the amino acid sequence <SEQ ID 1498>. This protein is predicted to be gp157. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3148(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD44102 GB:AF115103 orf157 gp [Streptococcus thermophilus

bacteriophage Sfi21]

-563-

Identities = 59/160 (36%), Positives = 100/160 (61%), Gaps = 3/160 (1%)

Query: 1 MAYLYELEGIYAQLQSMDLDEETFDQTLDSIDFQSDLENNIEYFVKMLKNVQADAKEYKA 60
 MA LYEL G + ++ +M++D+ET DTL++ID+ SD EN +E +VK++K+++AD E K
 5 Sbjct: 1 MATLYELTGQFLEIYNMEIDDETKLDTLEAIDWTSYENKVEGYVVKVKSLEADIEARKN 60

Query: 61 EKEAFYKKQKQAEAKAEKYKETIRLAMELSQKKKVDAGMFKVSLRRSKKVEILDETKIPL 120
 EK+ K ++K +K K + ++M + + +VD +FK+ +SK V +++E K+P
 10 Sbjct: 61 EKKRLDGLNKSDQSKIDKLKAALAI SMTETGQTRVDTTLFKIGFHKSKAV-VVNEEKLPK 119

Query: 121 DYMQEKEIEYKPMKAETSKALKSGIDISGVELIETESLQVK 160
 +Y + YKP K + + LKSG I G L E +L ++
 Sbjct: 120 EY--QIATYKPDKKTLLKELLKSGKHIEGATLEERRNLNIR 157

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 468

20 A DNA sequence (GBSx0505) was identified in *S.agalactiae* <SEQ ID 1499> which encodes the amino acid sequence <SEQ ID 1500>. This protein is predicted to be tropomyosin 2. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4474 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 469

35 A DNA sequence (GBSx0506) was identified in *S.agalactiae* <SEQ ID 1501> which encodes the amino acid sequence <SEQ ID 1502>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4114 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9649> which encodes amino acid sequence <SEQ ID 9650> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 470

A DNA sequence (GBSx0507) was identified in *S.agalactiae* <SEQ ID 1503> which encodes the amino acid sequence <SEQ ID 1504>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1505> which encodes the amino acid sequence <SEQ ID 1506>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3775(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 43/46 (93%), Positives = 46/46 (99%)

Query: 1 MTKQHRETLIWRASHQEREKLLDFGLVDKSYVTLRLRKKYAI 46
MTKQHRETLIWRASHQERE+LLDFGLVDK++YVTLRLRKKYAI
Sbjct: 1 MTKQHRETLIWRASHQERERLLDFGLVDKARYVTLRLRKKYAI 46

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 471

A DNA sequence (GBSx0508) was identified in *S.agalactiae* <SEQ ID 1507> which encodes the amino acid sequence <SEQ ID 1508>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1509> which encodes the amino acid sequence <SEQ ID 1510>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-565-

bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 76/77 (98%), Positives = 76/77 (98%)

Query: 1 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60
 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT
 10 Sbjct: 1 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60
 Query: 61 AEADRLRIERYKQENTL 77
 AEADRLRIERYKQEN L
 15 Sbjct: 61 AEADRLRIERYKQENAL 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 472

20 A DNA sequence (GBSx0509) was identified in *S.agalactiae* <SEQ ID 1511> which encodes the amino acid sequence <SEQ ID 1512>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1513> which encodes the amino acid sequence <SEQ ID 1514>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 52/127 (40%), Positives = 75/127 (58%), Gaps = 1/127 (0%)
 Query: 160 EDRFVDVVEANLGRGLVKFEFDMINDYLIGQNVSKDLFLEAVKVAVANNVRKPNYIARIL 219
 E + + + GR + FE + I ++ N+ ++ A++ AV NN + YI +IL
 45 Sbjct: 3 EKKLFEFNLQTFGRMISPFEIEDIQKWIHEDNMPIEVVNLALREAVENNKISWKYINKIL 62
 Query: 220 DNWINDGIKTPEQAYQAQRDFKAKKANKTMQSQSNNVPSWSNPDYKGPLKEFALGSIDDI 279
 +W G T E+ + F K +++ + SNVPSWSNPDYK PDL+EFALGS+D I
 Sbjct: 63 VDWYKSGDTTVEKVRDLQRFDSSKKQRSVTT-SNVPSWSNPDYKEPDLEEFALGSMDGI 121
 50 Query: 280 EDGSGDF 286
 EDGSGDF
 Sbjct: 122 EDGSGDF 128

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-566-

Example 473

A DNA sequence (GBSx0510) was identified in *S.agalactiae* <SEQ ID 1515> which encodes the amino acid sequence <SEQ ID 1516>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -5.63    Transmembrane    13 - 29 ( 11 - 31)

----- Final Results -----
10      bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9647> which encodes amino acid sequence <SEQ ID 9648> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 474

20 A DNA sequence (GBSx0511) was identified in *S.agalactiae* <SEQ ID 1517> which encodes the amino acid sequence <SEQ ID 1518>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5822(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 475

35 A DNA sequence (GBSx0512) was identified in *S.agalactiae* <SEQ ID 1519> which encodes the amino acid sequence <SEQ ID 1520>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-567-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 476

A DNA sequence (GBSx0513) was identified in *S.agalactiae* <SEQ ID 1521> which encodes the amino acid sequence <SEQ ID 1522>. This protein is predicted to be P1-antirepressor homolog. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3411(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9645> which encodes amino acid sequence <SEQ ID 9646> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG31333 GB:AF182207 ORF 169a [Bacteriophage mv4]
Identities = 88/167 (52%), Positives = 122/167 (72%)

Query: 100 MLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKKITNLTMMNNQLQLDLKEAQKQ 159
M+ + K K++R+YFIQVEK++NSPE I+ RAL +++ +I L +N L L L+E+ K+
Sbjct: 1 MMSKTAKGKEIRQYFIQVEKNWNNSPEMIQRAL EISNARIQELQAQNKSLTLQLEESNKK 60

Query: 160 ARYLDLIIESKGAIRVTQIAADYGMSVNFKNKTLLEFGVQHKVNGQWILYKRHMKGKGYTD 219
A YLD+I+ + L TQIAADYG S FN+ L E G+QHKVNGQWILYK +MGKGY
Sbjct: 61 ASYLDIILGTPDLLATTQIAADYGYSARTFNQLLKEVGIQHKVNGQWILYKAYMGKGYVQ 120

Query: 220 SHTFDYQDKNGHTRANVTTTWTQKGRFLYELLKDNNILPLIEQEDI 266
S +F ++D+ GH R+ +T WTQKGR +Y++LK+N LPLIE++DI
Sbjct: 121 SKSFAFKDRKGHDRSKPSTYWTQKGRKLIYDVLKENGTLPLIERDDI 167

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1523> which encodes the amino acid sequence <SEQ ID 1524>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4214(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/249 (52%), Positives = 163/249 (65%), Gaps = 14/249 (5%)

Query: 19 MNQLINITLNENQEPVVSGRDLHNVLNIKTQYTKWLERMSEYGFEEVDYIAISQKRLTA 78
MNQLIN+TLNENQEPVVSGRDLH VL IKTQYTKWLERMSEYGF EN D++AISQKRLTA
Sbjct: 1 MNQLINVTLNENQEPVVSGRDLHKVLEIKTQYTKWLERMSEYGFVENEDFMAISQKRLTA 60

Query: 79 QGNRT EYIDHVLKLDMAKEIAMLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKK 138
QGN+TEY DHVLKLDMAKEIAMLQRNEKSK+VRKYFIQVEKDFNSPEKIMARALLMADKK
Sbjct: 61 QGNQTEYTDHVLKLDMAKEIAMLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKK 120

Query: 139 ITNLTMMNNQLQLDLKEAQKQARYLDLIIESKGAIRVTQIAA-----DYGMSVNFKNKTL 193
+ ++L+ ++ + + + D + S ++ V ++A + + L
Sbjct: 121 V-----HKLEAQIEADRPKVLFAVASHTSILVGEALAKLLKQNGVNI GATRLFTWL 173

-568-

Query: 194 LEFGVQHKVNGQ-WIL-YKRHMKGKGYTDSHTFDYQDKNGHTRANVTTTWTQKGRLEFLYEL 251
 + G K NG+ W + ++ + G +GH + T T KG+ +
 Sbjct: 174 RKHGYLIKRNGRDWNMPTQKSVELGLIRVKETSITHSDGHITVSKTFLVTGKGQQYFINK 233

5 Query: 252 LKDNNILPL 260
 + LP+
 Sbjct: 234 FLNQEYLPV 242

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 477

A DNA sequence (GBSx0514) was identified in *S.agalactiae* <SEQ ID 1525> which encodes the amino acid sequence <SEQ ID 1526>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4205(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1527> which encodes the amino acid sequence <SEQ ID 1528>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 21/63 (33%), Positives = 31/63 (48%), Gaps = 1/63 (1%)

Query: 1 MQQFNLIKQLREKKGFTQNELADKANVSRSLVVGLETGSYSETSTASLKKLAKALDVKIKD 60
 M+ LK R K +Q LAD VSR + +E G Y+ T + + + LD + D
 Sbjct: 1 MKNLKLKAARAGKDLSQQALADLVGVSRQTIAAVEKGDYNPTINLCI-AICRVLDKTLDD 59

40 Query: 61 LFF 63
 LF+
 Sbjct: 60 LFW 62

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 478

A DNA sequence (GBSx0515) was identified in *S.agalactiae* <SEQ ID 1529> which encodes the amino acid sequence <SEQ ID 1530>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0396(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-569-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:BAA17582 GB:D90907 hypothetical protein [Synechocystis sp.]
 Identities = 45/164 (27%), Positives = 79/164 (47%), Gaps = 33/164 (20%)

Query: 102 EEEELRNLF TKLIASSMDKSKNEFNHPSFIEI IKQFDKIDAQNFKIISDLYFKKGFVATG 161
 ++E L+ L+ L+AS++ +S + SF+E++KQ D +DA+ ++ L+ +

10 Sbjct: 97 DDENLQTLWANLLASALTESDRTNSTKSFVEVLKQVDIVDAELNLVLYLLHLRV----- 150

Query: 162 TYVTTIIGQDKPLEHIASHVFVDNLEQNDIAIQSSSLTNLERLGLIQINY--KAHVDEKE 219
 KP E ++ D+ + N + I S +L NLERLGL+ I+ VDE+

Sbjct: 151 -----MAKPDEFTYAN---DSRKYNIVQI-SVALNNLERLGLLIHKKYDDTPVDEEA 198

15 Query: 220 YYNINLNSFITKKNSELKEQNKRVLTNLGMITLTLFGVRFSKTC 263
 +I ++ N K ++LTLFG+ F + C

Sbjct: 199 RISIW---YMQDGNRSFKAH-----VSLTLFGIHFMRVC 229

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1531> which encodes the amino acid
 sequence <SEQ ID 1532>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 64/215 (29%), Positives = 105/215 (48%), Gaps = 23/215 (10%)

Query: 65 QKLAKIEIQDVVSKNIE-NLQEPSLSIAGPALEASKFYLEEEELRNLF TKLIASSMDKSKN 123
 +K EI SK + +L+EP I PA+ S+ YL B LRN+F + IAS+ ++ K

35 Sbjct: 72 EKFKNEIDCEFSKIPQTSLEKPEVEYILYPAINESQYLSNETLRNMFARTIASTFNQDKE 131

Query: 124 EFNHPSFIEI IKQFDKIDAQNFKIISDLYFKKGFVATGTYYTTIIGQDKPLEHI----- 177
 + H +F++IIKQ +DAQN +I+ IG E++

Sbjct: 132 KDLHSAFVQIIKQMTPLDAQNILLINQ-----EGNNLIANLQIGVHYSKENLSGTVNK 184

40 Query: 178 ASHVFDNLEQNDIAIQSSSLTNLERLGLIQINYKAHVDEKEYYNINLNSFITKKNSELK 237
 A++++ L+ + I +SS+ NL RLGLI++Y + + Y +I + SE+

Sbjct: 185 ANNIYLSKLDYSPDII-ASSIDNLRGLIKVDYHLHYPLDSNYESIKQTTIYKSLESEIN 243

45 Query: 238 EQNKRVLTNL-----GMITLTLFGVRFSKTCL 264
 N +N G ++LT FG +F CL

Sbjct: 244 TLNLFKTSNTKYDIKIEKGKVSLSLDFGKKFISVCL 278

SEQ ID 1530 (GBS261) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 8; MW 31kDa).

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 479

A DNA sequence (GBSx0516) was identified in *S.agalactiae* <SEQ ID 1533> which encodes the amino acid sequence <SEQ ID 1534>. Analysis of this protein sequence reveals the following:

55 Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.55 Transmembrane 3 - 19 (1 - 26)

-570-

----- Final Results -----

5 bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 480

A DNA sequence (GBSx0517) was identified in *S.agalactiae* <SEQ ID 1535> which encodes the amino acid sequence <SEQ ID 1536>. Analysis of this protein sequence reveals the following:

15 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 35 - 51 (30 - 51)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1537> which encodes the amino acid sequence <SEQ ID 1538>. Analysis of this protein sequence reveals the following:

30 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.94 Transmembrane 31 - 47 (30 - 51)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 45/52 (86%), Positives = 48/52 (91%)
 Query: 1 MNWKKMLGDLHTFTSRDGKEKTSVEFEGGVLPALLVLLGGITWLIWLITK 52
 MNWKKLM GDLEHTFT+ DGKEKTS+EFEGGVLPALLVLLGGI W+IAW ITK
 40 Sbjct: 1 MNWKKLMFGDLHTFTTNHDGKEKTSIEFEGGVLPALLVLLGGIAWMIWFITK 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 481

45 A DNA sequence (GBSx0518) was identified in *S.agalactiae* <SEQ ID 1539> which encodes the amino acid sequence <SEQ ID 1540>. Analysis of this protein sequence reveals the following:

50 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>

-571-

```

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 482

- 10 A DNA sequence (GBSx0519) was identified in *S.agalactiae* <SEQ ID 1541> which encodes the amino acid sequence <SEQ ID 1542>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

```

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3934(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 483

- 25 A DNA sequence (GBSx0520) was identified in *S.agalactiae* <SEQ ID 1543> which encodes the amino acid sequence <SEQ ID 1544>. This protein is predicted to be repressor protein. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

- 30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0905(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 9643> which encodes amino acid sequence <SEQ ID 9644> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1545> which encodes the amino acid sequence <SEQ ID 1546>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 55
>>> Seems to have no N-terminal signal sequence
45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3117(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 175/264 (66%), Positives = 207/264 (78%), Gaps = 19/264 (7%)

-572-

Query: 34 LGKYIKKYRDTNNLSMAEFakesGISKAY--VSILEKNRDPNGKEIIPSIPIIKKVSdT 91
 LG I+K R+ N++ E ++ G+ K Y VS EKN + GK++ KK+++
 Sbjct: 24 LGDRIRKLREGNMTQTELSEILGM-KTYTTTVSKWEKNENFPKGKDL-----KKLAET 75
 5
 Query: 92 IGISFDDLNSLDENQIVALNETKTEKNLTSSTLQKITSTSSQLEQPRQEKVLSFANEQL 151
 ++ D LL L ++K K + +I S +QLEQPRQEKVL+FANEQL
 Sbjct: 76 FNVTS DYLLG-----LTDSKLGKITIQNEQPEIVSIYNQLEQPRQEKVLNFANEQL 126
 10
 Query: 152 EEQNKVVSFMFDRKVEETENYITDYVEGLVAAGLGAYQEDNLHMEVKLRADDVDPKYDTIA 211
 EEQNK VS+FD+K EETE+YITDYVEGLVAAGLGAYQEDNLHM+VKLR+DDVPD+YDTIA
 Sbjct: 127 EEQNKTVSIFDKKSEETEDYITDYVEGLVAAGLGAYQEDNLHMKVKLRSDVDPDEYDTIA 186
 15
 Query: 212 KVAGNSMEPLIQDNDLLFVKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLSLNKS 271
 KVAG+SMEPLIQDNDLLF+KVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLSLNKS
 Sbjct: 187 KVAGDSMEPLIQDNDLLFKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLSLNKS 246
 Query: 272 YEEIYLSENDNIRTIGEVDIYRE 295
 YEEIYLS++D+IRTIGEVDIYRE
 20 Sbjct: 247 YEEIYLSKDDIRTIGEVDIYRE 270

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 484

25 A DNA sequence (GBSx0521) was identified in *S.agalactiae* <SEQ ID 1547> which encodes the amino acid sequence <SEQ ID 1548>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3760 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 485

40 A DNA sequence (GBSx0522) was identified in *S.agalactiae* <SEQ ID 1549> which encodes the amino acid sequence <SEQ ID 1550>. This protein is predicted to be integrase (ripX). Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2719 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
 bacteriophage MM1]
 Identities = 36/59 (61%), Positives = 48/59 (81%), Gaps = 1/59 (1%)

-573-

Query: 2 KIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVMNKE-IINK 59
 KI + +H+FRHSHISFLAE G+P+ +IMDRVGHs+ K TL IYSHTT +M++ ++NK
 Sbjct: 312 KIEKNLSSHIFRSHISFLAESGLPIKSIMDRVGHsNAKMTLEIYSHTTEDMEDKLNVK 370

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1551> which encodes the amino acid sequence <SEQ ID 1552>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/71 (88%), Positives = 66/71 (92%)

Query: 1 MKIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVMNKEIINKQ 60
 +KIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVMNKEIINKQ
 Sbjct: 1 LKIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVMNKEIINKQ 60
 Query: 61 TAPFVPLLKSE 71
 T PF +K +
 Sbjct: 61 TDPFKTGIKQK 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 486

A DNA sequence (GBSx0523) was identified in *S.agalactiae* <SEQ ID 1553> which encodes the amino acid sequence <SEQ ID 1554>. This protein is predicted to be 50S ribosomal protein L19 (rplS). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3331(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9641> which encodes amino acid sequence <SEQ ID 9642> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC01534 GB:U88973 ribosomal protein L19 [Streptococcus thermophilus]
 Identities = 110/115 (95%), Positives = 112/115 (96%)
 Query: 25 MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 84
 MNPLIQSLTEGQLR+DIP FR GDTV RVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT
 Sbjct: 1 MNPLIQSLTEGQLRTDIPSFPRGDTV RVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 60
 Query: 85 VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRLALQGKAARIKEIR 139
 VRKIS GIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRLALQGKAARIKEIR+
 Sbjct: 61 VRKISSGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRLALQGKAARIKEIRK 115

-574-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1555> which encodes the amino acid sequence <SEQ ID 1556>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4849(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 111/115 (96%), Positives = 113/115 (97%)

Query: 25 MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 84
 MNPLIQSLTEGQLRSDIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT
 Sbjct: 1 MNPLIQSLTEGQLRSDIPNFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 60

Query: 85 VRKISGGIGVERTFPIHTPRVDKIEVVRVYGVRRRAKLYYLALQKGAARIKEIRR 139
 VRKISGGIGVERTFPIHTPRVDKIEV+R+GVRRRAKLYYLALQKGAARIKEIRR
 Sbjct: 61 VRKISGGIGVERTFPIHTPRVDKIEVIRHGKVRRAKLYYLALQKGAARIKEIRR 115

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 487

A DNA sequence (GBSx0524) was identified in *S.galactiae* <SEQ ID 1557> which encodes the amino acid sequence <SEQ ID 1558>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAC18596 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1 MKAQAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQGIMKMYSAQT 60
 MK QAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT
 Sbjct: 1 MKTQAIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQA KILADSGYQGLMKIYPQAQT 60

Query: 61 PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTFKIFSTTYRNRKRFGRLRMNLIA 120
 RKSSKLKPLT+EDK NH LSKER KVENIFAKVKTFK+PSTTYR+ RKRFGRLRMNL A
 Sbjct: 61 SRKSSKLKPLTVEDKACNHALSKERSKVENIFAKVKTFKMFSTTYRSHRKRFGRLRMNL A 120

Query: 121 GMINRELGF 129
 G+IN ELGF
 Sbjct: 121 GIINHELGF 129

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-575-

Example 488

A DNA sequence (GBSx0526) was identified in *S.agalactiae* <SEQ ID 1559> which encodes the amino acid sequence <SEQ ID 1560>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -10.99    Transmembrane    81 - 97 ( 67 - 107)
    INTEGRAL    Likelihood = -6.32     Transmembrane     8 - 24 ( 6 - 25)
    INTEGRAL    Likelihood = -2.76     Transmembrane    120 - 136 ( 120 - 136)

10 ----- Final Results -----
    bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database:
    >GP:BAB04382 GB:AP001509 unknown conserved protein in others
      [Bacillus halodurans]
      Identities = 53/150 (35%), Positives = 82/150 (54%), Gaps = 1/150 (0%)

20 Query: 1 MLNPYKRIFTLGLLATFLLFIFHFGRYSGLGTLNLIASFTNKNLYDYDWLLKLCITVITL 60
      M N R F GL+ L +I Y+G G +++E SFT +++ Y +L KL T +T+
      Sbjct: 251 MKNHTVRAVFVGLIIVALTYIIGSYDYNCRGLDMLDSFT-QDVPPYAFLAKLVFTAVTM 309

Query: 61 AAGYQGGEVTPLEAIGASLGVIITAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEV 120
      G+ GGE PLF +GA+LG + + LP+ +AALG FG NT + L+G E+
      Sbjct: 310 GMGFVGGEAIPLFVFGATLGNLTHAFIDLPLSFLAALGMIVTFGGGANTPIAAILLGVEV 369

Query: 121 FGFANTPYFVIVCLVAYSISHAHTIYGAQS 150
      F +F + CL +Y S H ++ +Q+
30 Sbjct: 370 FNGKGIEFFVACLTSYLFSGHGLWPSQT 399

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1561> which encodes the amino acid sequence <SEQ ID 1562>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
35 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.99    Transmembrane    56 - 72 ( 53 - 78)
    INTEGRAL    Likelihood = -7.17     Transmembrane    337 - 353 ( 327 - 355)
    INTEGRAL    Likelihood = -6.74     Transmembrane    264 - 280 ( 260 - 282)
    INTEGRAL    Likelihood = -6.16     Transmembrane    167 - 183 ( 161 - 187)
40 INTEGRAL    Likelihood = -5.26     Transmembrane    223 - 239 ( 217 - 242)
    INTEGRAL    Likelihood = -5.10     Transmembrane    20 - 36 ( 19 - 42)
    INTEGRAL    Likelihood = -0.37     Transmembrane    102 - 118 ( 102 - 119)
    INTEGRAL    Likelihood = -0.16     Transmembrane    300 - 316 ( 300 - 316)

45 ----- Final Results -----
    bacterial membrane --- Certainty=0.5798(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:
    >GP:BAB04382 GB:AP001509 unknown conserved protein in others
      [Bacillus halodurans]
      Identities = 129/397 (32%), Positives = 210/397 (52%), Gaps = 14/397 (3%)

55 Query: 20 VLGLVGLALPIGGAVGVVDVIFGKGLLFLSEYRDHHLFLLLPFLALAGLVIVFLYDKLG- 78
      +L + + IG VG + L E R++ + +L FL LAGL + +LY K G
      Sbjct: 9 LLTWIFFGIMIGAIVGSATALLTVNDHLGETRENRPWFVL-FLPLAGLALGYLYMKAGT 67

Query: 79 ---KEVRQGMGLVFQVGHGQKNQIPPMLIPLILFSTVWTHLFGASAGREGVAVQIGATIS 135
      E+ +G LV + G K ++ L PL+ T++T LFG S GREG A+Q+G +++
60 Sbjct: 68 SAGNELYKGNLVIESVQG-KGKMLLRGLPLVYLGTFTILFGGSTGREGAAIQMGGSVA 126

```

-576-

Query: 136 HYCRR-FVTSQEAARHLLIMGMAAGFAGLFQTPIAAVVFALEVLLVGTLRYSALLPSLVA 194
 + F R LL+ G++AGF F TPI A +F +E+ +G L++ AL+P LVA
 Sbjct: 127 EAVNKLKFKVLIDTRILLMGGISAGFGAAGFTPTAAIFGEMASLGRLLKFEALVPCLVA 186

Query: 195 AYVASWTSHALG-LEKFTIVLEBETLTITPLTLVKLIGLGLIFGLVGNSTFAYLL-GWFKPY 252
 ++V +T+ +E ++ ++ LT K+I L ++F LV + L G K
 Sbjct: 187 SFVGHYTTTEKFWHVEHEKFIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKLS 246

Query: 253 LSQKLPNPYPFRIAFIGALLSICL--MIGHVGRYSGLGTNLIAAAFSGQTILTIDWLLKMI 310
 + N R AF+G L+ + L +IG Y+G G +++ +F+ Q + Y +L K++
 Sbjct: 247 EKYTMKNHTVR-AFVGGLIIVALTYIIGSYD-YNGRGLDMLDSFT-QDVPPYAFLAKLV 303

Query: 311 VTVISLSAGFQGGVETPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPI 370
 T +++ GF GGE PLF +GA+LG L ++ LP+ +AALG FG NT A
 Sbjct: 304 FTAVTMGMGFVGGEAIPLFVFGATLNTLHAFIDLPLSFLAALGMIVTFGGGANPTPIAAF 363

Query: 371 FIGIEVFSPENALAYFVTSAAAYMVSHRHSIYSYQKV 407
 +G+E+F + +FV +Y+ S H ++ Q +
 Sbjct: 364 LLGVEMFNGKGIFFVACLTSYLSFGHHGLWPSQTI 400

An alignment of the GAS and GBS proteins is shown below:

Identities = 91/147 (61%), Positives = 111/147 (74%)

Query: 3 NPYKRIPTLGLLATFLLFIFHFGRYSGGLGTNLIEASFTNKNLYDYDWLLKLCITVITLAA 62
 NPY RI +G L + L I H GRYSGGLGTNL I A+F+ + + YDWLLK+ +TVI+L+A
 Sbjct: 259 NPYFRIAFIGALLSICLMIGHVGRYSGLGTNLIAAAFSGQTILTIDWLLKMIVTVISLSA 318

Query: 63 GYQGGVETPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEVFG 122
 G+QGGVETPLFAIGASLG+++AP LGLEPV+LVAALGYT+VFGSATNT PI IG EVFG
 Sbjct: 319 GFQGGVETPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPIFIGIEVFG 378

Query: 123 FANTPYFVIVCLVAYSISHAHTIYGAQ 149
 N + + AY +SH H+IY Q
 Sbjct: 379 PENALAYFVTSAAAYMVSHRHSIYSYQ 405

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 489

A DNA sequence (GBSx0527) was identified in *S.agalactiae* <SEQ ID 1563> which encodes the amino acid sequence <SEQ ID 1564>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.65	Transmembrane	47 - 63 (45 - 70)
INTEGRAL	Likelihood = -5.04	Transmembrane	219 - 235 (208 - 237)
INTEGRAL	Likelihood = -3.35	Transmembrane	168 - 184 (168 - 187)
INTEGRAL	Likelihood = -0.48	Transmembrane	141 - 157 (141 - 157)

----- Final Results -----

bacterial membrane	---	Certainty=0.4461(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9317> which encodes amino acid sequence <SEQ ID 9318> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04382 GB:AP001509 unknown conserved protein in others
 [Bacillus halodurans]
 Identities = 75/223 (33%), Positives = 119/223 (52%), Gaps = 18/223 (8%)

5

Query: 17 FSLLLIGGVVGAITAVFGRVLLFLTAFRSDYIAYLLPFLSIVGLFIVFVYQKFGGKS---- 72
F ++IG +VG+ TA+ V L R + ++L FL + GL + ++Y K G +
Sbjct: 15 FGIMIGAIVGSATALLLTVDNHLGETRENRPWFVL-FLPLAGLALGYLYMKAGTSAGNEL 73

10

Query: 73 VKGMGLVFEVHGEGNEETIPKRLVPLVILTTWLTHLFGGSAGREGVAVQIGATVSHYFQKY 132
KG LV E G + + RL PLV L T++T LFGGS GREG A+Q+G +V+ K
Sbjct: 74 YGNNLVIESVQKGKML-LRLGPLVYLGTFTILFGGSTGREGAAIQMGGSVAEAVNKL 132

15

Query: 133 CRLQNASQLFLVM-GMAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANF 191
+++ L+M G++AGF F TP+ A F +E+ +GRL ++P L+A+ ++
Sbjct: 133 FKVKLIDTRILLMGGISAGFGAAGFTPTTAAIFGEMASLGRCLKFEALVPCLVASFVGHY 192

Query: 192 VSHSLGLEKFHSH-----STATSMALTPDIILKLLVLGLCFGL 228
+ EKF H IAT ++ K+++L + F L
Sbjct: 193 TT-----EKFWHVEHEKFI IATVPEVSALTFKSSVILLAIIVPSL 230

A related GBS gene <SEQ ID 8577> and protein <SEQ ID 8578> were also identified. Analysis of this protein sequence reveals the following:

The protein has homology with the following sequences in the databases:

```

45      ORF01989(349 - 1491 of 1794)
       GP|4512350|dbj|BAA75315.1||AB011836(15 - 399 of 424) similar to Bordetella parapertussis
       transposase for insertion sequence element(27%-identity) {Bacillus halodurans}
       PIR|T44296|T44296 hypothetical protein [imported] - Bacillus halodurans
       %Match = 15.4
50      %Identity = 33.4   %Similarity = 54.7
       Matches = 129 Mismatches = 167 Conservative Sub.s = 82


       222          252           282            312             342              372               402                432
MY*RKSKTINLTM*YEQLSKTL*QNLVFIKKRIL*TVIKRFDNYAQYVLVLGFSLLIGGVVGAITAVFGRVILLFLTAFRS
                               |::||:||:||||:| || | |
                                MNKTFWLTLLTWIFFGMIGAIVGSATALLLVNDHLGETRE
                                      10                  20                   30                    40


       462          492           513            540             570              600               630                660
DYIAYLLPFLSIGLFIVFYVQKFG---GKSV-KGMGLVFVEVGHGNEETIPKRVLPLVILTITWLTHLFGGSAGREGVAVQ
 :    ::| || : || : ::| | | : || || | : : : || || | ||::| |||| | ||| |:|
NRPWFVL-FLPLAGLAGLYMKAGTSAGNELYKGNNLVIESVQG-KGKMLLRLGPLVYLGTFTMTILFCGGSTGREGAIIQ
                                   60                 70                  80                   90                     100                      110                       120

```

-578-

5

130 140 150 160 170 180 190 200

924 954 984 1014 1041 1071 1101 1131

EKFSHSIATSMALTPDIILKLLVLGLCFGLCGNLFAYLLAKA-KLIASSRLLNPNYKRIFTLGLLATFLLFIFHFGRYSGL

10

EHEKFIIATVPEVSALTFSKVILLAIIVFSLVSVLYCQLRHGIHKLSEKYTMKNHTVRAVFGGLIIVALTYIIGSYDYNCR

210 220 230 240 250 260 270 280

1161 1191 1221 1251 1281 1311 1341 1371

GTNLIEASFNTKNLYDYDWLLKLCITVITLAAGYQGGEVTPPLFAIGASLGVIIAPILGLPVIILVAALGYTSVFSGSATNTL

15

GGLDMLIEDSFT-QDVPPYAFLAKLVFTAVTMGMGFGGEAIPLFVFGATLGNTHAFIDLPFLSFLAALGMIVTFGGGANTP

290 300 310 320 330 340 350

1401 1431 1461 1491 1521 1551 1581 1611

LGPILIGGEVFGFANTPYFVIVCLVAYSISHAHTIYGQASR*LVMSFKRVYQFVERNIFFSFLFS*SL*KWSLSIL*MOK

20

IAAFLGLGVEMFNGKGIIEFFVACLTSYLFSGHHGLWFSQTIYEPKSRLYGVRKGETIKRTEEMKE

370 380 390 400 410 420

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 490

A DNA sequence (GBSx0528) was identified in *S.agalactiae* <SEQ ID 1565> which encodes the amino acid sequence <SEQ ID 1566>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:AAB98234 GB:U67480 chorismate mutase/prephenate dehydratase
(pheA) [Methanococcus jannaschii]
Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2 ELLEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDDNNRENEVLNNVLKKVQN 61
+L EIR++IDEID +++ L+ R L +V K + +P+ D RE + + + K +
Sbjct: 4 KLAETRKKIDEIDNKKLKLIAERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKE 63

45 Query: 62 HQFDDVIRATFKDIMTE-SRVYQKE 85
H D+ I I+ E ++ QK+
Sbjct: 64 HNWDENIGIKIFQILLIEHNKALQKQ 88

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1567> which encodes the amino acid sequence <SEQ ID 1568>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2356(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

60 An alignment of the GAS and GBS proteins is shown below:

-579-

Identities = 45/91 (49%), Positives = 62/91 (67%)

Query: 1 MELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDDNNRENEVLNNVLKKVQ 60

M LE+IRQEI+ ID LV+LLE RM L+ +V A+K + LPVLD REN++L+ V V+

Sbjct: 1 MRLEKIRQEINGIDHHLVALLEKRMALVEQVTAYKLANHLPVLDDQARENQILDRVSYLVK 60

Query: 61 NHQFDDVIRATFKDIMTESRVYQKENIVDGD 91

+ F+ I TFK IM+ SR YQ +++ GD

Sbjct: 61 DQAFEPAIHETFKTIMSLSRQVQTQHLTGGD 91

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 491

A DNA sequence (GBSx0529) was identified in *S.agalactiae* <SEQ ID 1569> which encodes the amino acid sequence <SEQ ID 1570>. This protein is predicted to be neuraminidase. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.35 Transmembrane 28 - 44 (28 - 47)

----- Final Results -----

bacterial membrane --- Certainty=0.2338(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10191> which encodes amino acid sequence <SEQ ID 10192> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA51473 GB:X72967 neuraminidase [Streptococcus pneumoniae]

Identities = 294/504 (58%), Positives = 380/504 (75%), Gaps = 10/504 (1%)

Query: 303 EDIKSYFYQYYCHLNHQLKLPKGAILSAKTEVYRGDFGRKNKDNVFGYRIPSLKTEKGT 362

E+++ Q + + + KLP+GA L+ KT+++ G G+ NKD + YRIP+LLKT+KGT

Sbjct: 299 EEVQKRSQLFKRSDLKLPKGAALTEKTDIFESGRNGKPNKDGKISYRIPALLKTDKGT 358

Query: 363 LLVGADERIEQACDWGNIGMVIRESDDGVITWCKRETTIVNLRNPRVPLVTSGDYSGSPI 422

L+ GADER + DWG+IGMVIRESDD+G TWG R TI NLR+NP+ S GSP+

Sbjct: 359 LTAGADERLHSSDWGDIGMVIRESDDNGKVTWCDRVTTITNLRDNPKA----SDPSIGSPV 414

Query: 423 NMDMALVQDTSSKIKRIFSIYDMFPEGKRGVSIANTPEKEYTQIGGQSYLNLYNNGKSK 482

N+DM LVQD +TKRIFSIYDMFPEG+G+ +++ E+ Y +I G++Y LY G+K

Sbjct: 415 NIDMVLVQDP--ETKRIFSIYDMFPEGKIGFMSSQKEEAYKKIDGKTYQILYREGEKG- 471

Query: 483 VFTIRDKGIVYNFKGKKTDYHVITETTKSDHSNLGDIYKKGQLLGNIFYFTKHKTSPFRLA 542

+TIR+ G VY GK TDY V+ + K +S+ GD+YKG QLLGNIFYFT +KTSPFR+A

Sbjct: 472 AYTIRENGTVYTPDGKATDYRVVVDPVKPAYSDKGDLYKGNQLLGNIFYFTTNKTSPPFRIA 531

Query: 543 KSSYVWMSYSDDDGRTWSSPRDITASLRQKGMKFLGIGPGKGIVLKWGPAGRIIIPAYS 602

K SY+WMSYSDDDG+TWS+P+DIT ++ MKFLG+GPG GIVL+ GPH GRI+IP Y+

Sbjct: 532 KDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYT 591

Query: 603 TNWKSHLRGSQSSRIIYSDDHGKTWHTGKAVNDNRILSNGEKIHSMTDNKKEQNTESVP 662

TN SHL GSQSSR+IYSDDHGKTWH G+AVNDNR + +G+KIHS TM+N++ QNTES

Sbjct: 592 TNNVSHLNGSQSSRIIYSDDHGKTWHAGEAVNDNRQV-DGQKIHSSTMNRRQNTTESTV 650

Query: 663 VQLKNGDIKLFMRNLGTGNLEVATSKDGGETWQNHVKRYKEVHDAYVQLSAIRFEHDKKEY 722

VQL NGD+KLFMR LTG+L+VATSKDGG TW+ +KRY +V D YVQ+SAI H+ KEY

Sbjct: 651 VQLNNGDVKLFMRGLTGLQVATSKDGGVTWEKDIKRYPVQKDVYVQMSAIHTMHEGKEY 710

-580-

Query: 723 ILLVNANGPGKKRQDGYARLAQVNRNGSFKWLYHHHIQDGSFAYNSVQQLNNDKFGVLYE 782
 I+L NA GP KR++G LA+V NG WL H+ IQ G FAYNS+Q+L N ++G+LYE
 Sbjct: 711 IILSNAGGP--KRENGMVHLARVEENGELTWLKHNPQKGEFAYNSLQELGNGEYGILYE 768

5 Query: 783 HREKHQNSFTLNKYVFNWSFLSQN 806
 H EK QN++TL+++ FNW FLS++
 Sbjct: 769 HTEKGQNAVTLSEFRKFNWDFLSKD 792

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 492

A DNA sequence (GBSx0530) was identified in *S.galactiae* <SEQ ID 1571> which encodes the amino acid sequence <SEQ ID 1572>. This protein is predicted to be unnamed protein product (gatC). Analysis of
 15 this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.63	Transmembrane	154 - 170 (149 - 178)
INTEGRAL	Likelihood = -11.99	Transmembrane	103 - 119 (98 - 123)
20 INTEGRAL	Likelihood = -7.91	Transmembrane	21 - 37 (14 - 40)
INTEGRAL	Likelihood = -6.53	Transmembrane	448 - 464 (444 - 467)
INTEGRAL	Likelihood = -5.89	Transmembrane	47 - 63 (45 - 68)
INTEGRAL	Likelihood = -5.10	Transmembrane	356 - 372 (352 - 373)
INTEGRAL	Likelihood = -4.78	Transmembrane	330 - 346 (328 - 350)
25 INTEGRAL	Likelihood = -4.41	Transmembrane	376 - 392 (375 - 393)
INTEGRAL	Likelihood = -3.72	Transmembrane	243 - 259 (235 - 266)
INTEGRAL	Likelihood = -2.55	Transmembrane	277 - 293 (275 - 293)

----- Final Results -----

30	bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1573> which encodes the amino acid
 35 sequence <SEQ ID 1574>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.31	Transmembrane	154 - 170 (150 - 179)
INTEGRAL	Likelihood = -11.68	Transmembrane	104 - 120 (99 - 124)
40 INTEGRAL	Likelihood = -9.82	Transmembrane	447 - 463 (442 - 469)
INTEGRAL	Likelihood = -7.91	Transmembrane	22 - 38 (11 - 41)
INTEGRAL	Likelihood = -7.11	Transmembrane	377 - 393 (375 - 403)
INTEGRAL	Likelihood = -5.89	Transmembrane	48 - 64 (46 - 69)
INTEGRAL	Likelihood = -4.78	Transmembrane	331 - 347 (329 - 351)
45 INTEGRAL	Likelihood = -3.88	Transmembrane	357 - 373 (353 - 373)
INTEGRAL	Likelihood = -2.55	Transmembrane	278 - 294 (276 - 294)
INTEGRAL	Likelihood = -1.22	Transmembrane	240 - 256 (240 - 257)

----- Final Results -----

50	bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 419/482 (86%), Positives = 447/482 (91%)

Query: 1 MQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAI 60
 MQ FL+I+NK I +GSGVVMLIVMTGLAMIFGVKF+KALEGGIKLAIALTGIGAI

-581-

Sbjct: 2 MQPFLDIINKILGFPIQLGSGVVMILIVMTGLAMIFGVKFTKALEGGIKLAIALTIGIGAI 61

Query: 61 GILTGAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPYTLYFLLIMLIVNIVMIVMK 120
GILTGAFSESLQAFVKNTGI+L+IIDVGWAPLATITWGSPYTLYFLL+ML+VNIVMIVMK

5 Sbjct: 62 GILTGAFSESLQAFVKNTGISLNIIDVGWAPLATITWGSPYTLYFLLVMLVNVNIVMIVMK 121

Query: 121 KTDTLDVDIFDIWHLSTIGLLIMWYAKNNLPTLLSVIIATVAIIFVGVVKIINSDLMKP 180
KTDTLDVDIFDIWHLSTIGLLIMWYA +N+LP +S++IATVA+I VGVVKIINSDLMKP

10 Sbjct: 122 KTDTLDVDIFDIWHLSTIGLLIMWYAARNHLPVFSLLIATVAVILVGVVKIINSDLMKP 181

Query: 181 TFDDLLGTGTPSMTSTHMNYMNPIMVLDKLFDPKVPGLDKYDFDAAKLNKAIGFWGS 240
TFDDLLGTGP SPMTSTHMNYMNPIMVLDK+FDKVPGLDKYDFDAAKLNK IGFWGS

Sbjct: 182 TFDDLLGTGPSPMTSTHMNYMNPIMVLDKIFDKVPGLDKYDFDAAKLNKKIGFWGS 241

15 Query: 241 KFFIGMILGLVIGIMGNPVFSFAALGGWFSLGFTAGACLELFSWFIAAVEPLSQGI 300
KFFIGM LG VIGIMG+P F+ ++ WF LGFTAGACLELFSWFIAAVEPLSQGI

Sbjct: 242 KFFIGMALGFVIGIMGDPHFTVESIKNWFLGFTAGACLELFSWFIAAVEPLSQGI 301

Query: 301 TNFANGKMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLVEAILLSKVGNGILPLAGII 360
TNFAN +MHGRRFNIGLDWPFIAGRAEIWACANILAPIML+EA+LLSKVGNGILPLAGII

20 Sbjct: 302 TNFANARMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLIEAVLLSKVGNGILPLAGII 361

Query: 361 AMGVTALLVVTGRRLIRMITFGTLLPLPFLSCTMIAPFATELAKKVGAFAPGARAGSL 420
AMG+TPALLVVTGRRLIRMI FG+LLLPLPFLSCTMIAPFATELAKKVGAFAPAG AGSL

25 Sbjct: 362 AMGMPALLVVTGRRLIRMIIFGSLLLPLPFLSCTMIAPFATELAKKVGAFAPAGTSAGSL 421

Query: 421 ITHSTLEGPMKIFGYVIGKATTGQLSAIIITLIIFATAYLGLFMWYAKQMKRRAEYAAAN 480
ITHSTLEGPMKIFGYVIG+ATTGQ++IITLIIF YL LF WYA QMK RNAEYA

30 Sbjct: 422 ITHSTLEGPMKIFGYVIGQATTGQIASIITLIIFVAIYLSLFAWYANQMKARNAEYAKT 481

Query: 481 QK 482
K

Sbjct: 482 MK 483

35 A related GBS gene <SEQ ID 8579> and protein <SEQ ID 8580> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
McG: Discrim Score: 4.31
GvH: Signal Score (-7.5): -2.64

40 Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 6 value: -12.63 threshold: 0.0

INTEGRAL	Likelihood = -12.63	Transmembrane	154 - 170 (149 - 178)
INTEGRAL	Likelihood = -11.99	Transmembrane	103 - 119 (98 - 123)
45 INTEGRAL	Likelihood = -7.91	Transmembrane	21 - 37 (14 - 40)
INTEGRAL	Likelihood = -5.89	Transmembrane	47 - 63 (45 - 68)
INTEGRAL	Likelihood = -4.88	Transmembrane	243 - 259 (235 - 265)
INTEGRAL	Likelihood = -1.22	Transmembrane	268 - 284 (268 - 284)
PERIPHERAL	Likelihood = 0.85	127	

50 modified ALOM score: 3.03

*** Reasoning Step: 3

----- Final Results -----

55 bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

60 ORF00838(343 - 1122 of 1455)
EGAD|91348|EC2092(9 - 344 of 451) PTS system, galactitol specific IIC component
{Escherichia coli} OMNI|NT01EC2494 PTS system galactitol-specific enzyme IIC component
SP|P37189|PTKC ECOLI PTS SYSTEM, GALACTITOL-SPECIFIC IIC COMPONENT (EIIC-GAT) (GALACTICOL-
PERMEASE IIC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).
65 GP|1736809|dbj|BAA15955.1||D90847 PTS system, Galactitol-specific IIC component (EIIC-GAT)

-582-

```

(Galactitol- permease IIC component) (Phosphotransferase enzyme II, C component).
{Escherichia coli} GP|17884
%Match = 10.9
%Identity = 29.8 %Similarity = 59.2
5 Matches = 68 Mismatches = 88 Conservative Sub.s = 67

282      312      342      372      402      432      462      492
LS*HI*NWN*S*RRRNMQVFLNIVNKKFFDPIIHMGSVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAIIGILT
10      |: :| |:: ||: :: |: |:| :: |: : | : || | :|::
      MFSEVMRYILDLGPTVMLFIVIIIFSKILGMKAGDCFKAGLHIGIGFVGIGLVIGLML
      10      20      30      40      50

522      552      582      612      642      672      702
GAFSESLQAFVNTGINSIIDVGWAPLATITWGSPTYLYFLIMLIVNIVMIVMKKTDITLDVIDFDIWHLSITGLLIM-
15      : : :| :| :|| :||| : :|| | | : | :||: |: : | :||:|:|:|:|:| :| :| :
DSIGPAAKAMAENFDNLHVVDVGWPGSSPMTWASQIALVAIPAILVNVVAMLITRMTRVVNVDIWNIWHMTFTGALLHL
      70      80      90      100      110      120      130

747      774      804      834      864      894
20 -----WYAKKN-NLPTLLSVIIATVAIIFVGVVKIINSDLMKPTFDDLLGTGPTSPMTSTH
      |:|: |: | : | :|
ATGSGWMIGMAGVVIHAAPVYKLGDFARDTRNFFLEGLIAPHGTSAYMG-----
      150      160      170      180

924      954      984      1014      1044
25 MNYMNPPIIMVLDKLFDPKVPGLDKYDFDAAKLNKAIGFWGSKFFIGMILGLVIXIM-----
      || :::| : :|: ||:: | | : : | :| :| :||:| :| :
-----PIAVLVDAIIEKI-PGVNRIKFSADDIQRKFGPFGEPTVGVFMGLIIGILAGYDVKGVLQAVKTAAVML~~~
      200      210      220      230      240      250

1092      1122      1152      1182      1212      1242
30 -----GNPVFSFASIRWLVFVFLQQGACLEVGLF*LVSWVQLLQ*NHFLRKLILLMVNAXX*
      || |: | : : || :
~VVSASLIFIPILTILIAVCVPGNQVLPFGDLATIGFVAMAVVHRGNLFRTLISGVIIIMSITLWIATQTIGLHTQLAAN
35      320      330      340      350      360      370      380

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 493

40 A DNA sequence (GBSx0531) was identified in *S.galactiae* <SEQ ID 1575> which encodes the amino acid sequence <SEQ ID 1576>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

45 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1577> which encodes the amino acid sequence <SEQ ID 1578>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

55 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

60 An alignment of the GAS and GBS proteins is shown below:

-583-

Identities = 85/100 (85%), Positives = 96/100 (96%)

Query: 1 MIKILAACGAGVNSSHQIKDAIETQLGDRGYNVHCDAMVVKDITEEMVNKYDIFTPIAKT 60
 MIKILAACGAGVNSSHQIKDAIETQ+ DRGY+VHCDAMVVKDITEE+V++YDIFTPIAKT
 5 Sbjct: 1 MIKILAACGAGVNSSHQIKDAIETQMSDRGYDVHCDAMVVKDITEELVSRDYDIFTPIAKT 60

Query: 61 DLGFNVPPIPVVEAGPILYRIPVMSEPVFTALEQVIKEHNL 100
 DLGF +PIP+VEAGPILYRIP+MSEPVF LE+VIKE++L
 10 Sbjct: 61 DLGFEMPIPIVEAGPILYRIPIMSEPVFAELERVIKENHL 100

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 494

A DNA sequence (GBSx0532) was identified in *S.agalactiae* <SEQ ID 1579> which encodes the amino acid sequence <SEQ ID 1580>. This protein is predicted to be GatA. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2078(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 10193> which encodes amino acid sequence <SEQ ID 10194> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GF:AAG09977 GB:AF248038 GatA [Streptococcus agalactiae]
 Identities = 39/135 (28%), Positives = 76/135 (55%), Gaps = 9/135 (6%)

30 Query: 16 QEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLDGTPFIQYVAIPHETQYVC 75
 Q L +++S+ L+++ YV + +A+ +RE+++PTGL+++ VAIPHT ++Y
 Sbjct: 20 QTNLLEVLISQYLLQKGYVKTEFSEKAILQREKDYPTGLQLE-----NMAVAIPHITYSEYV 73

35 Query: 76 LVDRIYFVKNSQPITFKHMINPEEECRVQDFFFIINSRN-SNQSDILSNLITFFITKGNL 134
 L I+ K +PI+F M E+E + + ++ N +Q+ +L+ L+T F +
 Sbjct: 74 LKPFYIYNKLKEPISFIQM-GTEDEIVMARYVIVLGISNPKDQAGLLAELMTLFSNPKIV 132

40 Query: 135 DRLHELGDNKEKINH 149
 +L E+ KE + +
 Sbjct: 133 QQL-EMAQTKEALKN 146

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1581> which encodes the amino acid sequence <SEQ ID 1582>. Analysis of this protein sequence reveals the following:

45 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3130(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 102/154 (66%), Positives = 122/154 (78%)

55 Query: 4 VTQDILFIDAHSQEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLDGTPFIQ 63

-584-

V +ILF +A +Q ELFDLV+ L K YV+ Y QA+ ERE FPTGLK+DLKDG+ I
 Sbjet: 1 VFPNILFTEARTQPELFDLVASHLEKVGYYVTQFYHQALVEREAVFPTGLKVDLKDGSIL 60
 Query: 64 YVAIPHTEQYCLVDRIFYVKNQSPITPKHMINPEEECRVQDFFFIINSRNSNQSDILSN 123
 Y AIPHTET+YCLVD++ YV+NSQ +TFKHMINPEE+C V DFFFIINS+N Q+ ILSN
 Sbjet: 61 YAAIPHTEQYCLVDQVYVVRNSQALTFKHMINPEEDCLVTDFFFFIINSQNEGQTILSN 120
 Query: 124 LITFFITKGNLDRHLHELGNKEKINHYLIEKGVF 157
 LITFFITKGNL L L D+K+ I++YLIEKGVF
 Sbjet: 121 LITFFITKGNLSYLA SLKDDKQALSNY LIEKGVF 154

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 495

15 A DNA sequence (GBSx0533) was identified in *S.agalactiae* <SEQ ID 1583> which encodes the amino acid sequence <SEQ ID 1584>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1429(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA25176 GB:M60447 repressor protein [Lactococcus lactis]
 Identities = 139/255 (54%), Positives = 189/255 (73%), Gaps = 6/255 (2%)

30 Query: 1 MLKRERLQKIIKVNINGIVTVNEIMEELDVSDMTVRRDLDELKAGLLIRIHGGAQKVN 60
 M K+ RL+KI++ + I+G +T+ EI++ELD+SDMT RRDLD L+ GLL R HGGAQ ++
 Sbjet: 7 MNKKRRLEKILDMLKIDGTITTIKEIIDELDISDMTARRDLDALEADGLLTRTHGGAQLLS 66

Query: 61 ASPTPQNYEKSNTKEYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLD--- 117
 + + EK++ EK + T EK++IA+ A I DG+TIFIGPGTTL +LA +L
 35 Sbjet: 67 SK---KPLEKTHIEKKS LNTKEKIDIAKKACSLIKGDTIFIGPGTTLVQLALELKGRKG 123

Query: 118 FKIRVVTNSLPVFNILNQSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSN 177
 +KIRV+TNSLPVF ILN S T+DL+L+GGEYREITGAFVGS+ ++K++ F+KAFV +N
 40 Sbjet: 124 YKIRVITNSLPVFLILNDSSETIDLLLGGEYREITGAFVGSMASTNLKAMRFAKAFVRAN 183

Query: 178 GVFEKSIATYDEGEGETQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNI 237
 V SIATY + EG IQ++ALMN+ EKFLLVDS KF +YDF+ FY LD +D ++TD+ I
 Sbjet: 184 AVTHNSIATYSKKEGVIQQLALNNAVEKFLLVDSKFDYDFNFYNLQDLDTIITDNQI 243

45 Query: 238 DNVVKEQYSSFTKIL 252
 E++S +T IL
 Sbjet: 244 SPQHLEEFQYTTIL 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1585> which encodes the amino acid
 50 sequence <SEQ ID 1586>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0740(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-585-

Identities = 161/252 (63%), Positives = 195/252 (76%), Gaps = 3/252 (1%)

Query: 1 MLKRERLQKIIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELKAGLLIRIHGGAQKVN 60
 MLKRERL KI E VN GIVTVN+I++ L+VSDMTVRRDLDEL+KAG LIRIHGGAQ +
 5 Sbjct: 1 MLKRERLLKITEIVNEQGIVTVNDIIQTLNVSDMTVRRDLDELEKAGKLIRIHGGAQSIT 60

Query: 61 ASPTFQNYEKSNTTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLDFKI 120
 P E+SN EK +QT EK E+A +A Q +NDGETIFIGPGTTLE A QL + +I
 10 Sbjct: 61 M---PNKKERSNIEKQTVQTKKEKWEASYATQLVNDGETIFIGPGTTLECFAEQLKNRQI 117

Query: 121 RVVNSLPLVFNILNQSSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSNGVF 180
 R+VNSLPLVFNIL S T+DLIL+GGEYR ITGAFVGS+ +I SL F+KAF+S NG++
 Sbjct: 118 RIVNSLPLVFNILQDSETIDLILIGGEYRSITGAFVGSLSQNISSLKFAKAFISCNGIY 177

Query: 181 EKSIATYDEGEGEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNIDNV 240
 + IATY E EGEIQ++A NNS EK+LLVD+OKF YDF+ FY L++ID V+TD I
 15 Sbjct: 178 KNDIATYSETEGEIQKLAFNNSIEKYLLVDNQKFNAYDFFIFYHLNNIDAVVTDTSQITED 237

Query: 241 VKEQYSSFTKIL 252
 V E+YS FT++L
 20 Sbjct: 238 VIERYSQFTQLL 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 496

A DNA sequence (GBSx0534) was identified in *S. agalactiae* <SEQ ID 1587> which encodes the amino acid sequence <SEQ ID 1588>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
 Identities = 86/371 (23%), Positives = 136/371 (36%), Gaps = 79/371 (21%)

40 Query: 11 DLSESELKAAQEFSLSGKSEANQDKPKTGKTAQEIEAIEPKAIVKPEDLLFGIAQATDYK 70
 DL++ + L K D TG IEP+ V L AT
 Sbjct: 526 DLTQIAFAEQELMLKDKKHRYRIDVDTG-----IEPLAVDVSSLPMHAGNATYDT 576

45 Query: 71 NGTFVIPHDKDHYHYVELKWFDEEKDLLADSDKTSLEDYLATAKYMMHPEKRPKVEGWG 130
 +FVIPH DH H V W + +AT KY M HPE RP V W
 Sbjct: 577 GSSFVIPHIDHIVVPYSLWRNQ-----IATIKYVMQHPEVRPDV--WS 619

50 Query: 131 KDAEIIYKEKDSNKADKPSAPPTDNKSTSNSSDKNLSSAEVFKQAKPEKIVPLDKIAAHMA 190
 K + + + P+ P D ++ + SA EV +K + + AA
 Sbjct: 620 KPGH-----EESGSVIPNVTPLDKRAGMPNWQIHSABEV-----QKALAEGRFAA--- 665

55 Query: 191 YAVGFEDDQLIVPHHDHYHNPMAWFDKGGELWKAPEGYTLQQLFST--IKYMEHPNPEL 248
 D I D W D +G +L+ + + + EL
 Sbjct: 666 -----PDGYIFDPRDLAKETFWWKDGSFISPRADGSSSLRTINKSDLSQAEWQQAQELL 719

60 Query: 249 KEKGWGHDSHDHNGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSDDEEP-EKQED 307
 +K G +D +K P+E+ + +K ++ ++P E ++
 Sbjct: 720 AKKNAGDATDTDK-----PEEKQQ-----ADKSNENQQPSEASKE 754

Query: 308 ESELDEYELGMAQNAKKYGMDRQSFQKLIQLSNKYSVSFESFNINGSQVQVTKKDGSK 367
 E E D++ + YG+DR + E + QL+ K ++ + VQ K+G
 Sbjct: 755 EKESDDF----IDSLPDYGLDRATLEDHINQLAQKANID-PKYLIFQPEGVQVFNKNGEL 809

-586-

Query: 368 VLVDIKTLTEV 378
 V DIKTL ++
 Sbjct: 810 VTYDIKTLQOI 820

A related DNA sequence was identified in *S.agalactiae* <SEQ ID 6983> which encodes the amino acid sequence <SEQ ID 6984>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8581> and protein <SEQ ID 8582> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 2
 McG: Discrim Score: 6.06
 GvH: Signal Score (-7.5): -5.61
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 0 value: 2.23 threshold: 0.0
 PERIPHERAL Likelihood = 2.23 6
 modified ALOM score: -0.95

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1589> which encodes the amino acid sequence <SEQ ID 1590>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 808/825 (97%), Positives = 816/825 (97%), Gaps = 3/825 (0%)

Query: 2 KKTGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYIDDSKGVKAPKTNKTMQ 60
 KKTGYIGSVAAILLATHIGSYQLGKHHMG ATKDNQIAYIDDSKGVKAPKTNKTMQ
 Sbjct: 2 KKTGYIGSVAAILLATHIGSYQLGKHHMSATKDNQIAYIDDSKGVKAPKTNKTMQ 60

Query: 61 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNYHFKQSDV 120
 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNY FKQSDV
 Sbjct: 61 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNYRKFQSDV 120

Query: 121 INEILDGYVIKVNNGNYVYLKPGSKRKNIRTKQOIAEQVAKGTKEAKEKGLAQVAHLSKE 180
 INEILDGYVIKVNNGNYVYLKPGSKRKNIRTKQOIAEQVAKGTKEAKEKGLAQVAHLSKE
 Sbjct: 121 INEILDGYVIKVNNGNYVYLKPGSKRKNIRTKQOIAEQVAKGTKEAKEKGLAQVAHLSKE 180

Query: 181 EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLDGDAYLVPHGNHYHYIPKDLSPSELAQAQ 240
 EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLDGDAYLVPHGNHYHYIPKDLSPSELAQAQ
 Sbjct: 181 EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLDGDAYLVPHGNHYHYIPKDLSPSELAQAQ 240

-587-

Query: 241 AYWSQKQGRGARPSPDYRPTFAP--GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPDAS 298
 AYWSQKQGRGARPSPDYRPTFAP GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPDAS
 Sbjct: 241 AYWSQKQGRGARPSPDYRPTFAPAPGRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPDAS 300

Query: 299 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAGYVVPBGDHYH 358
 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAGYVVPBGDHYH
 Sbjct: 301 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAGYVVPBGDHYH 360

Query: 359 IIPRSQLSPLEMEELADRYLAGQTDNDSGSDHSPSDKEVTHTFLGHRKAYGKGLDGKP 418
 IIPRSQLSPLEMEELADRYLAGQT+D+DSGSDHSPSDKEVTHTFLGHRKAYGKGLDGKP
 Sbjct: 361 IIPRSQLSPLEMEELADRYLAGQTEDDDSGSDHSPSDKEVTHTFLGHRKAYGKGLDGKP 420

Query: 419 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELV 478
 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADEL
 Sbjct: 421 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELA 480

Query: 479 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG YIMPKDGKDYFYARYQLDLTQIAFAEQELM 538
 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG YIMPKDGKDYFYAR QLDLTQIAFAEQELM
 Sbjct: 481 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG YIMPKDGKDYFYARDQLDLTQIAFAEQELM 540

Query: 539 LKDKKHYRYDIVDTGIEPRLAVDLSSLPMHAGNATYDTGSSSFVIPHIDHIHVVPYSWLTR 598
 LKDKKHYRYDIVDTGIEPRLAVD+SSLP MHAGNATYDTGSSSFVIPHIDHIHVVPYSWLTR
 Sbjct: 541 LKDKKHYRYDIVDTGIEPRLAVDVSSLP MHAGNATYDTGSSSFVIPHIDHIHVVPYSWLTR 600

Query: 599 NQIATIKYVMQHPEVRPDVWSKPGHEESG SVIPNVTPLDKRAGMPNWOI IHSABEVQKAL 658
 +QIATIKYVMQHPEVRPD+WSKPGHEESG SVIPNVTPLDKRAGMPNWOI IHSABEVQKAL
 Sbjct: 601 DQIATIKYVMQHPEVRPDIWSKPGHEESG SVIPNVTPLDKRAGMPNWOI IHSABEVQKAL 660

Query: 659 AEGRFAPDGYIFDPRDLAKETFWVKDGSFSIPRAGSSSLRTINKSDLSQAEWQQAQEL 718
 AEGRFA PDGYIFDPRDLAKETFWVKDGSFSIPRAGSSSLRTINKSDLSQAEWQQAQEL
 Sbjct: 661 AEGRFATPDGYIFDPRDLAKETFWVKDGSFSIPRAGSSSLRTINKSDLSQAEWQQAQEL 720

Query: 719 LAKKNAGDATDTDKPEEKQADKSNENQQPSEASK-EEKESDDFIDSLPDYGLDRATLED 777
 LAKKNAGDATDTDKP+EKQQADKSNENQQPSEASK EEKESDDFIDSLPDYGLDRATLED
 Sbjct: 721 LAKKNAGDATDTDKPEEKQADKSNENQQPSEASK EEKESDDFIDSLPDYGLDRATLED 780

Query: 778 HINQLAQKANIDPKYLIFQPEGVQFYNGELVTVYDIKTLQQINP 822
 HINQLAQKANIDPKYLIFQPEGVQFYNGELVTVYDIKTLQQINP
 Sbjct: 781 HINQLAQKANIDPKYLIFQPEGVQFYNGELVTVYDIKTLQQINP 825

SEQ ID 8582 was expressed in *E.coli* in two different forms. GBS293dNterm was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 14; MW 74kDa + lanes 17 & 18; MW 48.8kDa). GBS293C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figures 148 (lane 2-4; MW 71kDa + lanes 5 & 7; MW 46kDa) and 182 (lane 7; MW 46kDa). Purified GBS293C-His is shown in Figure 241, lanes 8 & 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 497

A DNA sequence (GBSx0535) was identified in *S.agalactiae* <SEQ ID 1591> which encodes the amino acid sequence <SEQ ID 1592>. Analysis of this protein sequence reveals the following:

possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-588-

The protein has homology with the following sequences in the GENPEPT database:

```

>GF:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
Identities = 213/463 (46%), Positives = 277/463 (59%), Gaps = 41/463 (8%)

5   Query: 4   KKT V-I I S A L S V A L F G T G V G A Y Q L G S Y N A --Q K S D N S V S Y V K T D K S D S K A Q A T A V N K T P D 60
      Sbjct: 2   K K T Y G I G S V A A I L L A T H I G S Y Q L G K H M G L A T K D N Q I A Y I --D D S K G K V K A P K T N K T M D 59

10  Query: 61   Q I S K E E G I S A E Q I V V K I T D D G Y V T S H G D H Y H Y N G K V P Y D A I I S E E L I M K D P S Y V F N K A D 120
      Sbjct: 60   Q I S E E G I S A E Q I V V K I T D G Y V T S H G D H Y H + Y N G K V P Y D A I I S E E L + M D P + Y F + + D 119

      Query: 121 V I N E V K D G Y I I K V N G K Y I L Y L K E G S K R T N V R T K E Q I Q K R E E W S K G G S K G E S G K H S S A K T 180
      Sbjct: 120 V I N E + D G Y + I K V N G Y Y + Y L K G S K R N + R T K + Q I + Q + + K E + + A + 174

15  Query: 181 Q A L S ----A S V R E A K A S G R Y T T D D G Y V F S P T D V I D D M G D A F L V P H G D H F H Y I P K A D L S P S 236
      Sbjct: 175 A H L S K E E V A A V N E A K R Q G R Y T T D D G Y I F S P T D I I D D L G D A Y L V P H G N H Y I P K K D L S P S 234

20  Query: 237 E L S A A Q A Y W N R K T G R S G N S S --K P S N S S S Y I H A S A P S G N V S T G R H A N A P I S I P R V T H A N H 294
      Sbjct: 235 E L A A A Q A Y W S Q K Q G R G A R P S D Y R P T P A P G R R K A P I P D V T P N P G Q G H Q P D -----N G G Y H 288

25  Query: 295 W S K P A G N H A T A P K H A P T T K P I N K D S A L D K M L K R L Y A Q P L Y A R H V E S D G L V Y D P A Q V N A F 354
      Sbjct: 289 P A P P R N D A S Q N K H Q ----R D E F K G K T F K E L L D Q L H R L D L K Y R H V E E D G L I F E P T Q V I K S 344

30  Query: 355 T A I G V S I P H G N H F H F I H Y K D M S P L E L E -A T R M V A E H R G H H I D A L G K K D S T E K P K H I S H E P 413
      Sbjct: 345 N A F G Y V V P H G D H Y H I I P R S Q L S P L E M E L A D R Y L A -----G Q T D D N D S G S D H S K P S 394

      Query: 414 N K E -P H T E E H H A V T P K D Q R K G K P ---N S Q I V Y S A Q E T E E A K K 452
      Sbjct: 395 D K E V T H T F L G H R I K A Y G K G L D G K P Y D T S D A Y V F S K E S I H S V D K 437

```

There is also homology to SEQ ID 1590.

SEQ ID 1592 (GBS94) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 3; MW 52.5kDa).

40 GBS94-His was purified as shown in Figure 194, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 498

45 A DNA sequence (GBSx0536) was identified in *S. agalactiae* <SEQ ID 1593> which encodes the amino acid sequence <SEQ ID 1594>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> May be a lipoprotein

```

```

50  ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

55 There is also homology to SEQ IDs 1596 and 5548.

-589-

A related GBS gene <SEQ ID 8583> and protein <SEQ ID 8584> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: 22   Crend: 5
      McG: Discrim Score:      13.64
      GvH: Signal Score (-7.5): -5.75
          Possible site: 24
      >>> May be a lipoprotein
      ALOM program   count: 0 value:  4.83 threshold:  0.0
10      PERIPHERAL Likelihood =  4.83      33
      modified ALOM score: -1.47

      ----- Final Results -----
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

SEQ ID 8584 (GBS22) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 6; MW 35kDa).

20 The GBS22-His fusion product was purified (Figure 94A; see also Figure 193, lane 4) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 94B), FACS (Figure 94C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 8584 (GBS22) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 7 & 8; MW 35kDa).

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 499

A DNA sequence (GBSx0537) was identified in *S. agalactiae* <SEQ ID 1597> which encodes the amino acid sequence <SEQ ID 1598>. Analysis of this protein sequence reveals the following:

```

30      Possible site: 39
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL   Likelihood = -0.59   Transmembrane   19 - 35 ( 19 - 35)

35      ----- Final Results -----
          bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40      >GP:CAAS1352 GB:X72832 ORF1 [Streptococcus equisimilis]
          Identities = 104/145 (71%), Positives = 126/145 (86%)

      Query: 1  MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAYAVRKITSMRIFSDDEG 60
          MK+++QRV +ASVSI+ + G+I +GL+LLVG+ P+D ED+AYAVRKI +MRIFSD +G
45      Sbjct: 1  MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPPDAAEDLAYAVRKIVNMRIFSDADG 60

      Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPFTGAADPVKANQFYDIFNQELANHSVSE 120
          KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPFTGAA P A+QFYD FN++LA+ V VE
50      Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPFTGAAPDMASQFYDRFNEQLADFVPVE 120

      Query: 121 TGQFGADMQVSLINDGPVTIILDTK 145
          G FGADMQVSLINDGPVTI+LDTK
      Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145

```

-590-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1599> which encodes the amino acid sequence <SEQ ID 1600>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1430(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 103/145 (71%), Positives = 124/145 (85%)

Query: 1 MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAVAVRKITSMRIFSDDEG 60
 MK+++QRV +ASVSI+ + G+I +GL+LLVG+ P+D ED+AYAVRKI +MRIFSD +G
 Sbjct: 1 MKLVLQRVKEASVSIDCKIAGAINQGLLLLVGVGPDNAEDLAYAVRKIVNMRIFSDADG 60

Query: 61 KMNLISQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
 KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P A+Q YD FN++LA V VE
 Sbjct: 61 KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDLASQLYDSFNEQLAEFVPVE 120

Query: 121 TGQFGADMQVSLINDGPVTIIVLDTK 145
 G FGADMQVSLINDGPVTI+LDTK
 Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145

SEQ ID 1598 (GBS368) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 4; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 4; MW 45kDa).

GBS368-GST was purified as shown in Figure 215, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 500

A DNA sequence (GBSx0538) was identified in *S.agalactiae* <SEQ ID 1601> which encodes the amino acid sequence <SEQ ID 1602>. This protein is predicted to be stringent response-like protein (rel) (relA). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.32 Transmembrane 60 - 76 (60 - 76)

----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA51353 GB:X72832 stringent response-like protein
 [Streptococcus equisimilis]
 Identities = 647/739 (87%), Positives = 696/739 (93%), Gaps = 1/739 (0%)

Query: 1 MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
 M KEINLTGEEVVA+ ++YM+ETD AFVK AL+YATAAH+YQ RKSGEFYI+HPIQVAGI
 Sbjct: 1 MAKEINLTGEEVVALAAKYMNETDAAFVKALDYATAAHFYQVRKSGEFYIVHPIQVAGI 60

-591-

Query: 61 LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQ 120
 LADLHLDAVTVACGFLHDVVEDT+ITLD IE DFGKDVRDI+DGVTKLGKVEYKSHEEQ
 Sbjct: 61 LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDGKDVRDIVDGVTKLGKVEYKSHEEQ 120

5 Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
 Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

10 Query: 181 SRIKWELEDLSFRYLNTEFEYKISHMMSEKRREREELVDIIVDKIRSYTEEQGLYGDYIG 240
 SRIKWELEDL+FRYLNTEFEYKISHMM+EKRRERE LVD IV KI+SYT EQGL+GD+YG
 Sbjct: 181 SRIKWELEDLAFRYLNTEFEYKISHMMNEKRREREALVDDIVTKIKSYTTEQGLFGDVYIG 240

15 Query: 241 RPKHIYSIYRKMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
 RPKHIYSIYRKMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
 Sbjct: 241 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

20 Query: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAE+GVAHWAYKKG+ KVNQAEQ V
 Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

25 Query: 361 GMDWIELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
 GM WI+ELVELQD S DA DFVDSVKEDIF+ERIVFTP GAVQELP++SGPIDFAYAI
 Sbjct: 361 GMDWIELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

30 Query: 420 HTQVGEEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
 HTQVGEEKA GAKVNGRMVPLTAKLKTGDVVEI+TNPNFSGPSRDWIK+VKNKARNKIRQ
 Sbjct: 421 HTQVGEEKAIGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKLVKTNKARNKIRQ 480

35 Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEILPRVSVKSEELYAAGVFG 539
 FFKNQDKE S+NKGR++LV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAGVFG
 Sbjct: 481 FFKNQDKELSVNKGRLMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAGVFG 540

40 Query: 540 DLSPIISIFNKLTEKERREERAKAKAEADELINGGEIKTDKRDVLKVSSENGVLIQAGSG 599
 D+SP+S+FNKLTEKERREERAKAKAEA+EL+NGGEIK + +DVLKV+SENGVLIQAGSG
 Sbjct: 541 DISPVSVFNKLTEKERREERAKAKAEAEELVNGGEIKHENKVDLVKVSSENGVLIQAGSG 600

45 Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDGSKKE 659
 LLMRIAKCCNPVPGD IEGYITKGRG+AIHR+DC N+KSQ+ Y++RLI+VEWD D S K+
 Sbjct: 601 LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLNSSKD 660

50 Query: 660 YMAEIDIYGLNRSGLLNDVLQTLNATKLVSTVNAOPTKDMKFANIHVSFGISNLAQLTT 719
 Y AEIDIYGLNR GLLNDVLQ LSN+TK +STVNAOPTKDMKFANIHVSFGI NL LTT
 Sbjct: 661 YQAEIDIYGLNRRGLLNDVLQILSNSTKSISTVNAOPTKDMKFANIHVSFGIPNLTHLTT 720

55 Query: 720 VVDKIKIIPDVYSVKRTNG 738
 VV+KIK +PDVYSVKRTNG
 Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1603> which encodes the amino acid
 sequence <SEQ ID 1604>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 64 - 80 (64 - 80)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

>GP:CAA51353 GB:X72832 stringent response-like protein
 [Streptococcus equisimilis]
 Identities = 700/739 (94%), Positives = 721/739 (96%)

65 Query: 5 MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64

		MAK +N+TGEEV+ALAA YM + D AFV KAL YATAAHFYQVRKSGEPYIVHPIQVAGI	
	Sbjct: 1	MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPYIVHPIQVAGI	60
5	Query: 65	LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQ	124
		LADLHLDAVTVACGFLHDVVEDTDITLD IE DFG D RDIVDGVTKLG+VEYKSHEEQ	
	Sbjct: 61	LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDGKDVDRDIVDGVTKLGKVEYKSHEEQ	120
	Query: 125	AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	184
		AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	
10	Sbjct: 121	AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	180
	Query: 185	SRIKWELEDLAFRYLNETEFYKISHMMKEKRREREALVEAIVSKVKTYYTQQLFGDVY	244
		SRIKWELEDLAFRYLNETEFYKISHMM EKRREREALV+ IV+K+K+YTT+QGLFGDVY	
15	Sbjct: 181	SRIKWELEDLAFRYLNETEFYKISHMMNEKRREREALVDDIVTKISYTTQQLFGDVY	240
	Query: 245	RPKHIYSIYRMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI	304
		RPKHIYSIYRMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI	
	Sbjct: 241	RPKHIYSIYRMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI	300
20	Query: 305	AAPKANGYQSIHTTVYGFPGPIEQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV	364
		AAPKANGYQSIHTTVYGFPGPIEQIRTK+MHQVAEYGVAAHWAYKKGVRGKVNQAEQ V	
	Sbjct: 301	AAPKANGYQSIHTTVYGFPGPIEQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV	360
25	Query: 365	GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI	424
		GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPK+SGPIDFAYAI	
	Sbjct: 361	GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI	420
	Query: 425	HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ	484
		HTQ+GEKA GAKVNGRMVPLTAKLKTGDVVEI+TN NSFGPSRDW+KLVKTNKARNKIRQ	
30	Sbjct: 421	HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNNSFGPSRDWIKLVKTNKARNKIRQ	480
	Query: 485	FFKNQDKELSVNKGRLDLSVYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVCGF	544
		FFKNQDKELSVNKGRLDLSVYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVCGF	
35	Sbjct: 481	FFKNQDKELSVNKGRLDLSVYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVCGF	540
	Query: 545	DISPISVFNKLTEKERREERAKAKAEAEELVKGGEVKHENKDLKVRSENGVIIQGASG	604
		DISP+SVFNKLTEKERREERAKAKAEAEELV GGE+KHENKDLKVRSENGVIIQGASG	
	Sbjct: 541	DISPVSVFNKLTEKERREERAKAKAEAEELVNGGEIKHENKDLKVRSENGVIIQGASG	600
40	Query: 605	LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLNSSKD	664
		LLMRIAKCCNPVPGDPI+GYITKGRGIAIHR+DC+NIKSQDGYQERLIEVEWDLNSSKD	
	Sbjct: 601	LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLNSSKD	660
45	Query: 665	YQAEIDIYGLNRGGLNDVQLILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT	724
		YQAEIDIYGLNR GLLNDVQLILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT	
	Sbjct: 661	YQAEIDIYGLNRGGLNDVQLILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT	720
	Query: 725	VVEKIKAVPDVYSVKRTNG	743
		VVEKIKAVPDVYSVKRTNG	
50	Sbjct: 721	VVEKIKAVPDVYSVKRTNG	739

An alignment of the GAS and GBS proteins is shown below:

Identities = 635/739 (85%), Positives = 691/739 (92%), Gaps = 1/739 (0%)

55	Query: 1	MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI	60
		M K +N+TGEEV+A+ + YM++ DVAFV AL YATAAH+YQ RKSGEPII+HPIQVAGI	
	Sbjct: 5	MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI	64
60	Query: 61	LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVDRDIIDGVTKLGKVEYKSHEEQ	120
		LADLHLDAVTVACGFLHDVVEDT+ITLDEIE DFG D RDI+DGVTKLG+VEYKSHEEQ	
	Sbjct: 65	LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQ	124
	Query: 121	AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	180
		AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	
65	Sbjct: 125	AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI	184
	Query: 181	SRIKWELEDLSFRYLNETEFYKISHMMSEKRREREALVDIIVDKIRSYTEEQGLYGDYI	240

-593-

SRIKWELEDL+FRYLNTEFYKISHMM EKRERE LV+ IV K+++YT +QGL+GD+YG
 Sbjct: 185 SRIKWELEDLAFRYLNTEFYKISHMMKEKREREALVEAIVSKVTYTTQQGLFGDVYG 244

Query: 241 RPKHIYSIYRMRDKKRFDQIYDLIAIRCIMETASDVYAMVGVIHELWRPMEGRFKDYI 300
 RPKHIYSIYRMRDKKRFDQI+DLIAIRC+MET SDVYAMVGVIHELWRPMEGRFKDYI
 Sbjct: 245 RPKHIYSIYRMRDKKRFDQIFDLIAIRCVMETQSDVYAMVGVIHELWRPMEGRFKDYI 304

Query: 301 AAPKANGYQSIHTTVYGPKEPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
 AAPKANGYQSIHTTVYGPKEPIEIQIRTK+MHQVAE+GVAHWAYKKG+ KVNQAEQ+V
 Sbjct: 305 AAPKANGYQSIHTTVYGPKEPIEIQIRTKDMHQVAEYGVAAHWAYKKGVGRKVNQAEQAV 364

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGFIDFAYAI 419
 GM WI+ELVELQD S DA DFVDSVKEDIF+ERIYVFTP GAVQELP+ESGFIDFAYAI
 Sbjct: 365 GMNWIQELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGFIDFAYAI 424

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
 HTQ+GEKATGAKVNGRMVPLTAKLKTGDVVEIITN NSFGPSRDW+K+VKTNKARNKIRQ
 Sbjct: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEILPRVSVKSEEALYAAVGFG 539
 FFKNQDKE S+NKGR+LLV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
 Sbjct: 485 FFKNQDKELSVNKGRLDLYSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544

Query: 540 DLSPISIFNKLTEKERREERAKAKAEADELINGGEIKTDKRDVLKVSSENGVIIQGASG 599
 D+SPIS+FNKLTEKERREERAKAKAEA+EL+ GGE+K + +DVLKV+SENGVIIQGASG
 Sbjct: 545 DISPISVFNKLTEKERREERAKAKAEAEELVKGGEVKHENKDVVKVSENGVIIQGASG 604

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDGSKKE 659
 LLMRIAKCCNPVPGD I+GYITKGRG+AIHRSDC N+KSQ+ Y++RLI+VEWD D S K+
 Sbjct: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLNSSKD 664

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
 Y AEIDIYGLNRSGLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL LTT
 Sbjct: 665 YQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724

Query: 720 VVDKIKIIPDVYSVKRTNG 738
 VV+KIK +PDVYSVKRTNG
 Sbjct: 725 VVEKIKAVPDVYSVKRTNG 743

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 501

A DNA sequence (GBSx0539) was identified in *S. agalactiae* <SEQ ID 1605> which encodes the amino acid sequence <SEQ ID 1606>. This protein is predicted to be 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (cpdB). Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.79 Transmembrane 779 - 795 (778 - 797)

----- Final Results -----
 bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12613 GB:Z99108 similar to 2',3'-cyclic-nucleotide
 2'-phosphodiesterase [Bacillus subtilis]
 Identities = 297/630 (47%), Positives = 419/630 (66%), Gaps = 21/630 (3%)

Query: 102 KVDLRIMSTDLHTNLVNYDYQDKESQKIGLAKTAVLIEAKKENPNTVLVDNGDVIQG 161
 +V L I++TTD+H N+++YDYY DKE+ GLA+TA LI++ +++NPNT+LVDNGD+IQG

Sbjct: 42 QVHLSILATTDIHANMMDDYDYSDKETADFGFLARTAQLIQKHREQNPNTLLVDNGDLIQG 101

Query: 162 TPLGTYKAIVKP---VAENEEHPMYQAMNALGYDASTLGNHEFNGLDYLLKKIATANLP 218
PLG Y + ++ + HP+ MNAL YDA TLGNHEFNGLD+L I A+ P

5 Sbjct: 102 NPLGEXAVKYQKDDIISGTGKTHPIISVMNALKYDAGTLGNHEFNGLDFLDGTIKGADFP 161

Query: 219 ILNANVLDFKTHQPVFKTYDIITKTFKSTGRAVALNIGITGIVPPQILNWDKANLEGKV 278
I+NANV + + + Y I KT D G + +G G VPPQI+ WDK NLEG+V

10 Sbjct: 162 IVNANVKT-TSGENRYTPYVINEKTLIDENGNEQKVKGVIYGFVPPQIMTWKKNLEGQV 220

Query: 279 IVKDSVKAIEEIVPTMRAGADVILVLSHSGIGDDRYEEGEENVGYQIAS-IGVDVAVVT 337
V+D V++ E +P M+A+GADVI+ L+H+GI G EN + +A+ KG+DA+++

Sbjct: 221 QVQDIVESANETIPKMKAEAGADVIALAHTGIEKQAQSSGAENAVFDLATKTKGIDAIIS 280

15 Query: 338 GHSHAEFPGNGTGFYEKYTGVDGIN---GKINGTPVTMAGKYGDHLGIIDLGLSYTNGK 394
GH H FPS +Y GV N G ING PV M +G +LG+IDL L +G

Sbjct: 281 GHQHGLFPSA-----EYAGVAQFNVEKGTINGIPVMPSSWGKYLGVIDLKLEKADGS 333

20 Query: 395 WQVSESSAKIRKIDMNSTTADERIALAKEAHDGTINYVRQQVGTITAPITSYFALVKDD 454
W+V++S I I N T+ +E + ++ H T+ YVR+ VG T A I S+FA VKDD

Sbjct: 334 WKVADSKGSIESIAGNVTSRNETVTNTIQQTHQNTLEYVRKPVGKTEADINSFFAQVKDD 393

Query: 455 PSVQIVNNAORWYVANELKGTPEANLPLLSAAAPFKAGTRGDATAYTDIPAGPVAIKNVA 514
PS+QIV +AQ+WY E+K T NLP+LSA APPKAG R A YT+IPAG +AIKNV

25 Sbjct: 394 PSIQIVTDAQWYAEKEMKDTYKKNLPLSAGAPFKAGGRNGANYYTNI PAGDLAIKNVG 453

Query: 515 DLVLYDNTVALLKVTGADLREWLEMSAGQFNQIDPNNAKAPQNIINTEYRTYNFDDVIDGLT 574
DLVLYDN ++K+TG++++WLEMSAGQFNQIDP Q ++N +R+YNFDDVIDG+T

30 Sbjct: 454 DLVLYDNTVQIVKLITGSEVKDWLEMSAGQFNQIDPAKGGDQALLNENFRSYNFDVIDGVT 513

Query: 575 YKFDITQPNKYNDGKVVNSQASRVRDLMYNGKPVADKQEFMIVTNNYRASGTFPGAKNA 634
Y+ D+T+P KYN++GKV+N+ +SR+ +L Y GKP++ QEF++VTNNYRASG G +

Sbjct: 514 YQVDVTKPAKYNENGVINADSSRIINLSYEGKPI SPSQEFVLVVTNNYRASGG-GGFPHL 572

35 Query: 635 TMNRLLN---LENRQTIINYIISEKTINPTADNNWGFTESIKDLRLRQFQADKAKNLVTN 691
T +++++ +ENRQ +++YII +KT+NP ADNNW + +L F+++ AK

Sbjct: 573 TSDKIVHGS AVENRQVLMQYII EQKT VNP KADNNWSIA-PVSGTNLTFESSLLAKPFADK 631

40 Query: 692 SKDIQYIASSTKDEGFGDYRFVYTEQEKVD 721
+ D+ Y+ S +EG+G Y+ + + D

Sbjct: 632 ADDVAYVGKSA-NEGYG VYKIQFDDDSNPD 660
Identities = 133/567 (23%), Positives = 214/567 (37%), Gaps = 147/567 (25%)

45 Query: 104 DLRIMSTTDLHTNLVNYDYQDKESQKIGLAKTAVLIEAKKENPNTVLVDNGDVIQGTP 163
DL +M T D H +L + A+ I E + E + +L+D GDV G

Sbjct: 668 DLTVMHTNDTHAHLDD-----AARRMTKINEVRSETNHNILLDAGDVFSGD- 713

Query: 164 LGTYKAIVKPAENEEHPMYQAMNALGYDASTLGNHEFNGLDYLLKKIATAN---LDYLKKIATAN--- 216
Y +A+ + MN +GYDA T GNHEF+ G D+L AT +

50 Sbjct: 714 --LYFTKWNGLAD-----LKMNMNMGYDAMTFGNHEFDKGPTVLSDFLSGNSATVDPAN 765

Query: 217 -----LEILNANVLDFKTHQPVFKTYDIITKTF----KSTGRAVALNIGITG--IV 262
PI++ANV +++P K++ +TF K G + + + G +

55 Sbjct: 766 RYHFEAPEFPIVSANV--DVSNEPKLSFVKKPQFTTAGEKKEAGIHPYILLDVDGEKVA 823

Query: 263 PPQILNWDKANLE--GKVIV-----KDSVKAIEEIVPTMRAGADVILVLSHSGIGD 312
+ D A GK IV +++VKAI+E + + I+ L+H G

Sbjct: 824 VFGLTTEDTATTSSPGKSIVFNDAFETAQNTVKAIQE-----EEKVNKIIALTHIG--- 874

60 Query: 313 DRYEEGEENVGYQIA-SIKGVDAVVTGSHAEFPGNGTGFYEKYTGVDGNGKINCTP- 370
N ++A +KG+D ++ GH+H T VD + N P

Sbjct: 875 -----HNRDLELAKVKVGIDLIIGHITH-----TLVDKMEVVNNNEPT 912

65 Query: 371 -VTMAGKYGDHLGIIDLGLSYTNGKQVSESSAKIRKIDMNSTTADERIALAKEAHDGT 429
V A +YG LG +D+ G Q +S+ + ID ++ E AK+ D

Sbjct: 913 IVAQAKEYGQFLGRVDVAFD-EKGVVQTDKSNLSVLPIDEHTEENPE-----AKQELDQF 966

Query: 430 INYV---RQQVGTITAPITSYFALVKDDPSVQIVNNAQRWYVANELKGTPEANLPLLSA 485

-595-

```

      N +      ++VG T      + + QR +V +      + A
Sbjct: 967 KNELEDVKNEKVGYT-----DVALDGGQREHVRTKEINLGNFIADGMLA 1009

Query: 486 AAPFKAGTRGDAT----AYTDIPAGPVAIKNVADLYLDNVTALLKVTGADLREWLEMSA 541
5      A  AG R  T      I  G + +  V ++ + N +  +TG ++E LE
Sbjct: 1010 KAKEAAGARIAITNGGGIRAGIDKGDITLGEVLNVMPFGNTLYVADLTGKQIKEALE--- 1066

Query: 542 GQFNQIDPNKAPQNIINTEYRTYNFDVIDGLTYKFDITQPNKYNKDGKVVNSQASRVRD 601
      Q + N E      F + G+ Y F +      NK G      +  V+
10 Sbjct: 1067 -----QGLSNVENGGGAFPOVAGIEYTFTLN-----NKPG----HRVLEVKI 1104

Query: 602 LMVNGKPVADKQE--FMIVTNNYRASG 626
      NG VA + + + TNN+ +G
15 Sbjct: 1105 ESPNGDKVAINTDDTYRVATNNFVGAG 1131

```

There is also homology to SEQ ID 1608. A related sequence was also identified in GAS <SEQ ID 9129> which encodes the amino acid sequence <SEQ ID 9130>. Analysis of this protein sequence reveals the following:

```

20 Possible cleavage site: 27

>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -4.67    Transmembrane    649 - 665 ( 648 - 666)
      INTEGRAL    Likelihood = -2.02    Transmembrane     6 - 22 (   5 - 22)
25 PERIPHERAL    Likelihood =  1.85

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8585> and protein <SEQ ID 8586> were also identified. Analysis of this protein sequence reveals the following:

```

30 Lipop: Possible site: -1    Crend: 7
    McG: Discrim Score:      6.68
    GvH: Signal Score (-7.5): 0.87
        Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.
35 ALOM program    count: 1 value: -5.79 threshold: 0.0
      INTEGRAL    Likelihood = -5.79    Transmembrane    779 - 795 ( 778 - 797)
      PERIPHERAL  Likelihood =  0.53      251
      modified ALOM score:  1.66

40 *** Reasoning Step: 3

      ----- Final Results -----
          bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

      LPXTG motif: 769-773

```

The protein has homology with the following sequences in the databases:

```

50 ORF01378(298 - 2337 of 3000)
    GP|6782402|emb|CAB70615.1||AJ133440(1 - 680 of 683) cyclo-nucleotide phosphodiesterase,
    putative {Strept
    ococcus dysgalactiae subsp. equisimilis}
    %Match = 38.3
55 %Identity = 59.0 %Similarity = 72.3
    Matches = 403 Mismatches = 181 Conservative Sub.s = 91

    105      135      165      195      225      255      285      315
    LFYHFLT*K*KKLEAQKELXTK*MCLTKLSFINKRLFLV*SLKIIRK*D*LNVFNKL**FL *DNIHVMF*WRRFMSKHY
60                                     |:| |
                                     MMTKGY

```

Example 502

Possible site: 24
>>> Seems to have no N-terminal signal sequence

-597-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0296(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 503

A DNA sequence (GBSx0541) was identified in *S.agalactiae* <SEQ ID 1611> which encodes the amino acid sequence <SEQ ID 1612>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1504(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 10195> which encodes amino acid sequence <SEQ ID 10196> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:CAB12860 GB:Z99109 similar to glucanase [Bacillus subtilis]
 Identities = 212/345 (61%), Positives = 268/345 (77%), Gaps = 1/345 (0%)
 Query: 30 SMETTLNYIKTLTSIPSPTGFTQTIMTYIIKELEAFGYSPIRTNKGGMVSLKGNNDTKH 89
 S+ T+ IK L SIPSPG T ++ YI L+ + +R +KGG++ +L G++ ++H
 Sbjct: 3 SVRKTMEIKELVSIPSPTGNTYEVINYIESLLKEWKVETVRNHKGGLIATLPGRDTSRH 62
 30 Query: 90 RMITAHLDTLGAMVRAIKPDGRLKIDLIGGYTYNAIEGENCTIHLKNGKEISGTALIHQ 149
 RM+TAH+DTLGAMV+ IK DGRLKIDLIGG+ YN+IEGE C I + +GK +GT L+HQ
 Sbjct: 63 RMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYQCIETA-SCKMYTGTILMHQ 121
 35 Query: 150 TSVHVKDAGTAERNQTNMEIRLDEKVTTADETRALGIQVGDFISFDPRITITDSGFIKS 209
 TSVHVKDAG AERNQ NMEIRLDE V +T LGI VGDF+SFDP R IT SGFIKS
 Sbjct: 122 TSVHVKDAGKAERNQENMEIRLDEPVHCRKDTTEELGIGVGDFVSDPRVEITSSGFIKS 181
 40 Query: 210 RYLDDKVSAGILMELLSVYKKEDIQLPYTTHFYFSAFEELGHGANSSIPNETVEYLAVDM 269
 R+LDDK S +L+ L+ + EDI+LPYTTHF S EE+G+G NS+IP ETVEYLAVDM
 Sbjct: 182 RHLDDKASVALLRLRIHEIQTEDIELPYTTHFLISNNEEIGYGNSNIPPETVEYLAVDM 241
 Query: 270 GAMGDDQETDEYTVSICVKDASGPYHYELRQHLVSLAENNNIPYKLDIYPYGSASAAAM 329
 GA+GD Q TDEY+VSICVKDASGPYHY+LR+HLV LAE ++I YKLDIYPYGSASAA+
 45 Sbjct: 242 GAIGDQATDEYSVSICVKDASGPYHYQLRKLHLVQLAEKHHIDYKLDIYPYGSASAAI 301
 Query: 330 RAGAEVKHALLGAGIESSHSYERTHIDSIQATELLVDAYLKSNNMV 374
 ++G ++ H L+G GI++SH++ERTH S++ T L+ Y++S MV
 50 Sbjct: 302 KSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV 346

There is also homology to SEQ ID 424.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 504

A DNA sequence (GBSx0542) was identified in *S.agalactiae* <SEQ ID 1613> which encodes the amino acid sequence <SEQ ID 1614>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3157(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF11472 GB:AE002031 conserved hypothetical protein [Deinococcus radiodurans]
Identities = 55/150 (36%), Positives = 85/150 (56%), Gaps = 2/150 (1%)
15
Query: 5  LIIIRGNSASGKSTIAKQLQAEIGENTLLLSQDYLRREMLGTKDGENTTTIPLLINLINY 64
      LI++RGNS SGKS++A+ L+  G  + QDYLR +L  D  I L+  + Y
Sbjct: 23  LIVLRGNSGSGKSSVARALRERFGYGLAWVEQDYLRRLREHDVAGGKNIGLIETNVRY 82

20
Query: 65  GYHNCYSIILEGILRSDWYTPVWKHILKHNPNNTYAYYYDLSPQETVVRHSTRKLSLEFG 124
      S +LEGIL S  Y P+ + + H  + +Y+DL F+ETV+RH+TR ++ +FG
Sbjct: 83  CLSAGSVTVLEGILFSRHYGPMLERL--HADFGGHWFYFDLPFEETVRRHATRPQAADFG 140

Query: 125 EDSLARWWLEKDFLKEIPEKILTKAMSLD 154
25     + W+ +D L + E+++ A SL D
Sbjct: 141 VQDMQAWFQARDVLPFVQEQLIGPASSLAD 170

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 505

A DNA sequence (GBSx0543) was identified in *S.agalactiae* <SEQ ID 1615> which encodes the amino acid sequence <SEQ ID 1616>. This protein is predicted to be periplasmic-iron-binding protein BitC. Analysis of this protein sequence reveals the following:

```

35 Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood =-11.46    Transmembrane    9 - 25 ( 5 - 30)

----- Final Results -----
40     bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45 >GP:AAD18094 GB:U75349 periplasmic-iron-binding protein BitA
      [Brachyspira hyodysenteriae] (ver 2)
      Identities = 114/331 (34%), Positives = 184/331 (55%), Gaps = 3/331 (0%)

Query: 11  YILLVVSIIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTPAFEEKY-GIKVKLI 69
      +I+ + ++ +++F  S SK  LVI  +  ++  + F+ K  I V+++
50 Sbjct: 4  FIIFCMLMSMTLFYSCSSGDSK--NANSLVIYCSHPLDLMTILDDFKAKNPDPINVEVV 61

Query: 70  QGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIPDYIHPSDTA 129
      GTG+L+ R+  E  D+ +GG  +S  LFE+Y S N  ++ ++ +
55 Sbjct: 62  TAGTGELLKRVEAEKMNPLGDVLWGGTLNSVKSCTDLFENYTSNEANILDEFKNTGPGF 121

```